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

by Jack Kleijnen and
Willem van Groenendaal

March 1994
Two-stage versus Sequential Sample-size Determination in Regression Analysis of Simulation Experiments

Jack P.C. Kleijnen and Willem Van Groenendaal

Tilburg University (Katholieke Universiteit Brabant), School of Management and Economics, CentER and 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 (simulation runs, replications) per combination of simulation inputs (system variant) 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 (meta)model. 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 stopping rule. The rules are first formalized and analyzed, and then evaluated through several Monte Carlo experiments. Both procedures turn out to yield confidence intervals for the estimated regression parameters that have the required 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)

1. Introduction

Regression analysis of simulation data gives metamodels, which are gaining increasing popularity. Nevertheless a number of research issues remain to be investigated, as Sargent (1991) witnesses. Welch (1990) also concludes that there is a deep need to investigate sample size selection for regression metamodels. Besides these research issues there is the following practical problem. Suppose the management scientists have built a simulation model of a queueing system. After running several system variants it turns out that one variant has a response with an estimated variance of 1 whereas another variant has a response with an estimated variance of 1,000. Wouldn’t management wonder why the simulation has been conducted with such wildly varying accuracies?

In this paper we concentrate on the assumption of constant response variance for the various combinations of simulation inputs. In practice, variances may indeed differ substantially;
for example, Kleijnen, Van den Burg and Van der Ham (1979, p. 60) report a simulation study of the Rotterdam harbor in which the 16 input combinations give estimated variances ranging between the values 64 and 93.228. (On the other hand, if the common variance assumption holds, the variance estimates may still vary substantially, because the variance estimators have large standard errors themselves; see the discussion leading to equation 5.5). When the variances seem not constant, then the analysts can accept this variance heterogeneity and use OLS (accounting for variance differences when deriving confidence intervals) or WLS; see Kleijnen, Cremers and Van Belle (1985). Such variance heterogeneity must be accepted in experiments with real-life systems and in socio-economic data analysis. In simulation, however, it is easy to select sample sizes (number of replications or runs) that reduce variance heterogeneity.

To select sample sizes, Sargent (1991) uses the guidelines in Whitt (1989). These guidelines, however, are heuristic formulas used to estimate required simulation run lengths in the early planning stages before any data have been collected. Moreover these formulas are entirely for queueing models.

A popular idea in the simulation field is to obtain more observations for those input combinations that have higher variability. Unfortunately, until now this idea was never formalized and analyzed in detail for metamodeling, although Welch (1990) sketches two approaches that resemble our two procedures. In this paper we analyze sequentialized (iterative, dynamic) sample size determination for any type of random simulation model. We focus on two stopping rules that are well-known in the statistics and the simulation literature: two-stage and sequential rules. We investigate whether the resulting coverage probabilities of the \((1 - \alpha)\) confidence intervals for the regression parameters are acceptable, that is, do the coverage probabilities not deviate significantly from the user specified value \((1 - \alpha)\)? Based on the literature we conjecture that two-stage sampling is 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 stopping rules is notoriously difficult. We shall show that an exact formalization in the context of 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 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 results give simula-
tion users some advice on how to apply the popular idea of taking more observations for noisy input combinations.

How much help does the literature provide for the issue under investigation? There is a long history of using two-stage and more general sequential sampling plans for achieving statistical precision. It is well known that this approach can be used in simulation to help determine an appropriate runlength in steady-state simulation or an appropriate number of replications in transient-state simulation. This is well demonstrated by the discussion of sequential procedures for steady-state means and of ranking and selection procedures in Law and Kelton (1991, pp. 563, 595-611). However, the statistical literature investigates stopping rules primarily in situations with only one or two populations (input combination); see the textbooks by Govindarajulu (1981) and Wetherill (1986) and the handbook edited by Ghosh and Sen (1991). Chaterverdi (1987) investigates sequential point estimation of regression parameters in a linear model, but he assumes a common variance and a specific family of loss functions. The simulation literature concentrates on sample-size selection for a single population; for a survey see Kleijnen (1987) and Kleijnen and Van Groenendaal (1992).

We use the following notation. We suppose that \( n \geq 2 \) input combinations are simulated, and that \( m_i \), the number of replications of combination \( i \) with \( i = 1, \ldots, n \) is determined sequentially. Sequentialization means that during the simulation experiment the responses \( y_{ij} \) \( (j = 1, \ldots, m_i) \) are obtained successively, and are used to estimate the \( n \) variances; these estimators are employed in a stopping rule. We underline random variables, because the random character of the number of replications is essential and becomes explicit in this notation. We further suppose that once the simulation experiment has stopped, its data are analyzed through the classic linear regression metamodel

\[
\vec{y} = X'\beta + \vec{e}
\]

(1.1)

with the vector of \( n \) normally distributed simulation responses \( \vec{y} = (y_1', \ldots, y_n')' \), the \( n \times Q \) matrix of independent variables \( X = (x_{iq}) \) with \( i = 1, \ldots, n \) and \( q = 1, \ldots, Q \), the vector of \( Q \) input effects or regression parameters \( \beta = (\beta_1, \ldots, \beta_Q)' \), and the vector of \( n \) fitting error \( \vec{e} = (e_1, \ldots, e_n)' \); we use bold upper case letters to denote matrices and bold lower case letters for vectors. Moreover, we assume that the metamodel has no specification error; that is, the fitting errors have zero expectation. The response variances vary with the inputs: \( \sigma_i^2 \neq \sigma^2 \). The responses are independent (no common pseudorandom number seeds).
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. In §3 we assume unknown variances and examine a two-stage sampling plan for regression metamodels. In stage #1 a pilot sample is taken to estimate these variances. Based on these estimates the rule computes how many more responses to simulate for each system variant. In §4 we present a sequential procedure, that is, after the pilot sample only one observation is taken from each combination that has a relatively large estimated variance for its sample average. In §5 we first specify a large Monte Carlo experiment. Next we analyze the results of this experiment. In §6 we summarize our results and indicate future research.

2. Known Response Variances

Only in this section we assume that the response variances $\sigma_i^2$ are known. Intuitively it seems wise to obtain more observations for those input combinations that have high variability (we shall return to this intuition). More specifically, suppose we take

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

(2.1)

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

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

(2.2)

The sample-size rule in (2.1) is not necessarily 'optimal', but it simplifies the regression analysis of the simulation data, as follows.

Note that there are several optimality criteria in the theory of optimal design; for example, observations can be taken for those input combinations that reduce the confidence intervals of interest most (also see Kleijnen 1987, p. 335). Welch (1990, p. 394) proposes to select the sample sizes such that the confidence intervals for $\beta_q$ will be of fixed widths; also see Chaturvedi (1987), who assumes a common variance and a specific family of loss functions. A rule closely related to (2.1) replaces the variances $\sigma_i^2$ by the standard errors $\sigma_i$; if the sample sizes were fixed and there were only two populations ($n=2$), then this alternative rule would minimize the variance of $\bar{y}_1 - \bar{y}_2$; see...

It is well-known that in case of known unequal variances the Best Linear Unbiased Estimator (BLUE) is provided by WLS applied to the individual responses \( y_{ij} \). It is simple to prove that this WLS estimator (say) \( \hat{\beta} \) follows from

\[
\min \sum_{i=1}^{n} \left\{ m_i (\hat{y}_i - \bar{y}_i)^2 / \sigma_i^2 \right\}
\]  

(2.3)

with the predicted responses \( \hat{y}_i = x_i' \hat{\beta} \) where \( x_i' \) denotes the \( i^{th} \) row of \( X \) so \( x_i' = (x_{ii}, \ldots, x_{iq}) \).

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 \( \sigma_i^2 \)). The OLS estimator \( \hat{\beta} \) is still unbiased (but not BLUE) and follows from

\[
\min \sum_{i=1}^{n} m_i \hat{y}_i - \bar{y}_i)^2.
\]  

(2.4)

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

\[
\min \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2.
\]  

(2.5)

In other words, sample-size rule (2.1) implies that \( \hat{\beta} \), 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 both conceptually and computationally, 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 \( \sigma_i^2 \) and hence \( m \), in eq. (2.1). Many sampling schemes are conceivable: two-stage, multistage, and sequential schemes. In this section we concentrate on a two-stage procedure (leading to so-called Stein estimators).

In stage #1 a pilot sample of \( m_0 \geq 2 \) responses is taken for each input combination. Then the \( n \) classic variance estimators are computed:
with the sample average of combination $i$ computed from $m_0$ simulation runs

$$
\bar{y}_i(m_0) = \frac{\sum_{j=1}^{m_0} y_{ij}}{m_0}.
$$

Because sample size is a crucial variable, we explicitly show the dependence of the various estimators on the sample size. Different values for the size of the pilot sample ($m_0$) 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

$$
m_i = \left[ m_0 s^2_i(m_0) / \min_{l \leq i \leq n} s^2_l(m_0) \right] .
$$

where $[x]$ denotes the integer closest to $x$.

So in stage #2, $m_i - m_0$ additional observations are taken from combination $i$ (so at the end of stage #2 there are $\bar{m}_i$ responses from combination $i$; at least one combination has exactly $m_0$ responses). All observations from a particular combination are used to estimate the mean and the variance of the response of that system:

$$
\bar{y}_i(\bar{m}_i) = \frac{\sum_{j=1}^{\bar{m}_i} y_{ij}}{\bar{m}_i},
$$

and

$$
s^2_i(\bar{m}_i) = \frac{\sum_{j=1}^{\bar{m}_i} (y_{ij} - \bar{y}_i(\bar{m}_i))^2}{(\bar{m}_i - 1)}. \tag{3.5}
$$

We shall need the following properties of these estimators for the $n$ means and $n$ variances respectively. The estimator of the expected simulation response (3.4) is the average of a random number of responses, that is, it is a ratio of two random variables. In general, such estimators are
biased. However, the estimator (3.4) remains unbiased, under our assumption of normally
distributed simulation responses. We prove this as follows. (i) The stage #1 statistics \( \bar{y}(m_i) \) and
\( s_i^2(m_i) \) are independent since the underlying populations are normal (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). (ii) Given \( s_i^2(m_i) = s_i^2(m_i) \), we have \( \bar{m}_i = m_i \) where \( m_i \) is a function of \( s_i^2(m_i) \) alone, and the stage #2 observations \( \{ y_{ij}: j = m_o + 1, \ldots, m_i \} \)
are i.i.d. with mean \( E(y_{ij}) \). Defining \( \bar{y}_i(\ast) = (\bar{m}_i - m_o)^{-1} \sum_{j=m_o+1}^{m_i} y_q \) it follows from (i) and (ii) that

\[
E(\bar{y}_i(\bar{m})) = E(E(\bar{y}_i(\bar{m} | s_i^2(m_i))))
\]

\[
= E(E(\{(m_o/\bar{m}_i) \bar{y}_i(m_o) + ((\bar{m}_i - m_o)/\bar{m}_i) \bar{y}_i(\ast) \} | s_i^2(m_i)))
\]

\[
= E(E(\{(m_o/\bar{m}_i) \bar{y}_i(m_o) | s_i^2(m_i)) + ((\bar{m}_i - m_o)/\bar{m}_i) \bar{y}_i(\ast) | s_i^2(m_i)))
\]

\[
= E(E(\{(m_o/\bar{m}_i) \bar{y}_i(m_o) + ((\bar{m}_i - m_o)/\bar{m}_i) \bar{y}_i(\ast) \} | y_{ij}))
\]

\[
= E(E(y_{ij})) = E(\{y_{ij}\}.
\]

However, the estimator of the variance of the sample average

\[
\text{var}(\bar{y}_i(\bar{m})) = \sum_{j=1}^{m_i} (y_q - \bar{y}_i(\bar{m}))^2 / \{(\bar{m}_i - 1)\bar{m}_i\}
\]

(3.7)
is biased, as we prove as follows. If we condition on all variance estimates \( s_i^2(m_i) \) -and hence
on \( \bar{m}_i \) - then we prove in appendix 1 that

\[
E(\{s_i^2(\bar{m}_i) | s_i^2(m_o) = s_i^2(m_o)\}) = s_i^2(m_o) + \{\sigma_i^2 - s_i^2(m_o)\}(\bar{m}_i - m_o)/\bar{m}_i = \sigma_i^2 + (s_i^2(m_o) - \sigma_i^2)m_o/\bar{m}_i.
\]

(3.8)

This equation means that if \( s_i^2(m_o) \) is less than \( \sigma_i^2 \), then the variance is underestimated; otherwise
the variance is overestimated. So the variance estimator in (3.7) is biased. 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 the estimators for the regression parameters are obtained. Equation (2.5) showed how to compute the WLS estimator \( \hat{\beta} \) for fixed sample sizes and known response variances. This equation inspires us to estimate the regression parameters in two-stage sampling through

\[ \hat{\beta}(\bar{m}) = (X'X)^{-1}X'y(\bar{m}) \]  

(3.9)

with \( \bar{m} = (m_0, m_1, ..., m_n)' \) the vector of \( n+1 \) sample sizes and \( y(\bar{m}) \) the vector of \( n \) sample means defined in (3.4). This regression estimator is unbiased because the response averages \( y(\bar{m}) \) are unbiased (see eq. 3.6) and we assume that the regression metamodel in (1.1) holds.

The covariance matrix of the WLS estimator in (3.9) is

\[ \text{cov}(\hat{\beta}(\bar{m})) = (X'X)^{-1}X'\text{cov}(\bar{y}(\bar{m}))(X'X)^{-1}, \]  

(3.10)

where \( \text{cov}(\bar{y}(\bar{m})) \) denotes the covariance matrix of the average responses at the end of stage #2. So to estimate \( \text{cov}(\hat{\beta}(\bar{m})) \) the user must estimate \( \text{cov}(\bar{y}(\bar{m})) \). Let us consider (i) the off-diagonal elements, and (ii) the diagonal elements:

(i) If the sample sizes were not random, then the covariances among the sample averages would be known to be zero (we assume the users do not employ common random numbers; see equation 1.1). Actually, the sample sizes are random; they might create dependence among the sample averages. Unfortunately, the users can not estimate the covariances among the sample averages, from a single simulation experiment. Therefore we assume these covariances to be zero. In our Monte Carlo experiment we can estimate these covariances (see equation 5.2); we shall find that this assumption is realistic indeed.

(ii) The users can compute a (biased) estimator of the variances of the sample means through (3.7).
In summary, we assume that the users employ (3.9) to estimate the regression parameters in the metamodel. This estimator resembles a WLS estimator: (3.9) does depend on the estimated weights (see 1/ω₂ in equation 2.3) because the sample sizes m do. To estimate the variances of the estimated regression parameters, the users employ the estimator \( \text{cov}(\hat{\beta}(\hat{m})) \), which is a diagonal matrix with elements defined by (3.7). Of course the users know X. Hence they can compute the estimator \( \text{cov}(\hat{\beta}(\hat{m})) \) as the sample analogue of (3.10).

The square roots of the diagonal elements of \( \text{cov}(\hat{\beta}(\hat{m})) \) 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 \( \beta_q \) we propose the classic Student statistic4. What is the correct number of degrees of freedom \( v \)? Scheffé (1970, p. 1502) proposes \( v = \min(m_i - 1) \) in his study of a case with only two populations \( (n = 2 \, : \, \text{Behrens-Fisher problem}) \); 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 \( m_i = m \); they investigate several degrees of freedom, namely \( \min(m_i - 1) \), \( \sum_{i=1}^n (m_i - 1) \), and infinity (\( \tau_v \) equals the 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 \( \min(m_i - 1) = m_i - 1 \) and infinity. We observe 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 degrees of freedom \( v \). This results in the \( 1 - \alpha \) confidence interval per regression parameter

\[
\hat{\beta}_q(\hat{m}) \pm t_{q}^{\alpha/2} \sqrt{\text{var}(\hat{\beta}_q(\hat{m}))} \quad \text{with } q = 1, ..., Q.
\]  

(3.11)

In the Monte Carlo experiment we shall investigate whether this interval covers the true value with prescribed probability of \( 1 - \alpha \). We observe that a joint confidence interval can be derived from the individual confidence intervals by applying Bonferroni's inequality (also see §5.3).

4. Sequential Sampling

The first stage of sequential sampling is identical to stage #1 of two-stage sampling: take a pilot sample of size \( m_0 \) from each of the \( n \) combinations and estimate the \( n \) response variances through (3.1). These variance estimators \( \hat{\sigma}_q^2(\hat{m}_0) \) are used to obtain the first estimators of the required number of replications \( \hat{m}_i \) through (3.3). Whereas two-stage sampling jumps ahead
and obtains \( \bar{m}_i - m_b \) observations, sequential sampling proceeds more cautiously: for each population with \( \bar{m}_i > m_b \) a single new response is obtained.

Next all estimated response variances are recomputed, except for the system with the smallest estimated response variance at the end of the pilot stage. To estimate the variance \( \sigma_i^2 \) at a particular stage (say ) t, all \( m_a \) responses from combination i that are available at that stage, are used. The total number of stages is not known beforehand; it is a random variable (say) \( T \) (so t runs from 1 through \( T \); for \( t = 1 \) we have \( m_n = m_b \) ). Hence we replace (3.5) by

\[
\sigma_i^2 = \frac{\sum_{j=1}^{m_a} (y_{ij} - \bar{y}_i(m_a))^2}{m_a - 1},
\]

(4.1)

where \( \bar{y}_i(m_a) \) denotes the average computed from the available \( m_a \) responses (analogous to eq. 3.4).

At the end of each stage the procedure re-estimates the required number of replications \( \bar{m}_i \), based on (3.3) but using the most recent variance estimate, defined in (4.1). The simulation continues until (for the first time) this desired number of replications does not exceed the available number \( m_n \), for all \( n \) combinations:

\[
\forall i : \bar{m}_i \leq m_n, \tag{4.2a}
\]

or equivalently:

\[
\forall i : s_i^2(m_a)/m_n \leq \min_{1 \leq i \leq n} s_i^2(m_a)/m_n. \tag{4.2b}
\]

So \( T \) denotes the minimum number of stages for which condition (4.2) holds; that is, response averages have essentially equal estimated variances (also see Welch 1990).

Note that we distinguish between \( m_a \), the number of available responses for combination i in stage t, and \( \bar{m}_a \), the estimated number of responses (for combination i in stage t) required to realize nearly constant response variances. To estimate \( \beta \) at the end of the sampling procedure, we assumed that the averages have constant variances; so we estimate \( E(y_{ij}) \) from \( \bar{m}_a \) responses. However, to estimate the nuisance parameters \( \sigma_i^2 \) during the various stages, we use as much information as possible, that is, we use \( m_a \) observations. In practice, the difference between \( \bar{m}_a \) and \( m_a \) is insignificant at the end of the sequential procedure when \( s_i^2(m_a) \) changes only slowly (also see Table 1). This component of our heuristic is partly supported by the following theory. In 1943 Scheffé devised a technique for the derivation of an exact confidence interval for the difference between the means of two normal popula-
tions with unequal variances (Behrens-Fisher problem); see Kleijnen (1974/1975, p. 473). Schef"f 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. 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 denote the maximum divided by the minimum estimated variance by \( r \). Then for the system with the smallest variance estimate we compute the response average from a single response \((m = 1)\); to estimate its variance, however, we use all responses for that system. The system with the largest estimated variance is simulated \( r \) times.

Once the procedure stops generating simulation responses, the metamodel is used to analyze the simulation data. Our WLS estimator based on a sequentially determined number of simulation responses is

\[
\hat{\beta}(\hat{m}_I) = (X'X)^{-1}X'\hat{y}(\hat{m}_I)
\]

with the vector of \((n + 1)\) sample sizes at the final stage \( \hat{m}_I = (m_0, \hat{m}_{I1}, \ldots, \hat{m}_{I_{m_I}})' \) and the vector of \( n \) sample means \( \bar{y}(\hat{m}_I) \) with components

\[
\bar{y}(\hat{m}_I) = \sum_{i=1}^{m_I} y_i / \hat{m}_{I_{i-1}}. \tag{4.4}
\]

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

The covariance matrix of this WLS estimator is

\[
\text{cov}(\hat{\beta}(\hat{m}_I)) = (X'X)^{-1}X'\text{cov}(\bar{y}(\hat{m}_I))X(X'X)^{-1}, \tag{4.5}
\]

where \( \text{cov}(\bar{y}(\hat{m}_I)) \) denotes the covariance matrix of the average responses at the end of stage \#I.

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

\[
\text{vár}(\bar{y}_{I_{m_I}}) = \hat{\sigma}^2(m_{I_{m_I}}) / \hat{m}_{I_{m_I}}. \tag{4.6}
\]

where the variance of an individual simulation response is estimated from all \( m_{I_{m_I}} \) available observations (see eq. 4.1), whereas the estimated variance of the average response accounts for
the fact that this average uses only \( \bar{m}_T \) responses (see the denominator of eq. 4.6). This estimator is biased; more specifically, this estimator underestimates the true variance: in case of sequential sampling (to be distinguished from two-stage sampling) it is well known that sampling until a lower threshold is reached for the estimated variance results in underestimation. This phenomenon is quantified in Welch (1990, figure 1). However, this bias may be negligible in large samples; see the Monte Carlo experiment in §5.

Our sequential procedure implies that the \( n \) variance estimates in (4.6) are virtually the same at the time the simulation stops: common estimated variances for the sample averages, at \( t = T \). So we assume that this procedure yields common true variances for \( \bar{y}(\bar{m}_T) \). We estimate this common variance by pooling these \( n \) variance estimators:

\[
\overline{s^2} = \frac{1}{n} \sum_{i=1}^{n} \text{var}(\bar{y}(\bar{m}_T)) .
\]  
(4.7)

But then the covariance matrix (4.5) reduces to a simple and familiar formula:

\[
\text{cov}(\bar{y}(\bar{m}_T)) = \overline{s^2} (X'X)^{-1} .
\]  
(4.8)

Note that \( X \) is often orthogonal in well-designed simulation experiments, as the Monte Carlo experiment in §5.1 demonstrates. Then the individual regression estimators become uncorrelated with constant estimated variances \( \overline{s^2}/n \). These properties, however, are not essential for our approach.

To obtain confidence intervals for \( \beta_q \), the Student statistic \( t_\nu \) is used, as in two-stage sampling. We again restrict ourselves to degrees of freedom \( \nu = \min(m_i - 1) = m_q - 1 \) and infinity respectively. This results in the \( 1 - \alpha \) confidence intervals of (3.11) where we replace \( \bar{m} \) by \( \bar{m}_T \). In the Monte Carlo experiment we shall investigate the coverage probabilities of these intervals; in Appendix 3 we illustrate sequential sampling through an example.

5. Monte Carlo Experiment

We discuss the design of our Monte Carlo experiment in §5.1, intermediate output in §5.2, and final output in §5.3.

5.1 Monte Carlo Inputs

The metamodel approach assumes that the linear regression model (1.1) treats the simulation model as a black box (approaches such as perturbation analysis and the score function method
do not see the simulation as a black box). So by definition, only the inputs and outputs of the simulation model are observed. For example, a queueing simulation may have one output (average waiting time) and six inputs (traffic rate, service rates of stations 1 through 4, and queuing priority); the network structure is not taken into account. We define the simulation inputs as those variables and statements of the simulation program that are changed during the simulation experiment. These inputs are called factors in experimental design literature. (Of course, the users' knowledge is essential, also in this black box approach: they must decide which factors are to be examined, over which domain the factors range, which transformations are to be tried for the factors, e.g., logarithmic transformations.)

So to study our sampling procedures we could have used the M/M/1 queue, as Sargent (1991) did, or some other more complicated queueing system. However, we would then have to rely on the asymptotic normality of the simulation responses. Moreover, the metamodel would have specification error; in other words, the fitting errors would not have zero expectation so the estimator of the regression effects \( \beta \) would be biased.

Therefore we conduct a Monte Carlo experiment with perfectly normal 'simulation' responses and with zero expectation fitting errors. In other words, we create a 'laboratory' that gives perfect statistical control over the experiment (moreover, it saves computer time); also see Kleijnen (1992, p. 1170).

In our Monte Carlo experiment we use nearly the same inputs as did Kleijnen, Cremers and Van Belle (1985), who studied fixed sample sizes. So we select the following three matrices of independent variables \( X \) in the metamodel (1.1).

(a) Consider again the Rotterdam harbor case study by Kleijnen et al. (1979). The users are interested in six simulation inputs (mean number of containers moved per ship, unbalance between loading and unloading of containers, arrival pattern, etc.). Users and analysts have studied the system before and assume that -besides the six main effects- there are interactions between specific factors, namely six two-factor interactions. So altogether the simulation experiment must estimate 13 effects, including the intercept (often denoted by \( \beta_0 \), but in this paper denoted by \( \beta_1 \)). The analysts are familiar with experimental design theory so they employ a \( 2^6-2 \) design to specify the 16 input combinations of the simulation.

(b) The second example is simpler: it concerns an artificial example with only three main effects besides the intercept. Hence \( X \) follows from a \( 2^3 \) design and is an \( 8 \times 4 \) matrix. (\( X \) is a submatrix of \( X \) in (a).)

(c) The third example is simplest: there are only two simulation inputs and these inputs
have no interaction. So $\mathbf{X}$ is determined by a $2^2$ design and is a $4 \times 3$ matrix. (This example follows from the examples in the appendixes 2 and 3.)

We combine these three $\mathbf{X}$ matrices with different degrees of variance heterogeneity for the simulation responses $\mathbf{y}$. In practical queueing simulations, variances may differ because traffic loads differ. We quantify variance heterogeneity by

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

(5.1)

For this $H$ we select the values 0, 4, and (roughly) 11. The individual response variances $\sigma_i^2$ are shown in appendix 4.

We assume that the users employ independent seeds for their pseudorandom number streams (see the text below equation 1.1). Hence $\mathbf{cov}(\mathbf{y})$, the true covariance matrix of the individual simulation responses, is diagonal.

The precise values of the true effects $\beta$ of the simulation inputs do not affect the interpretation of the Monte Carlo experiment; for example, whether the queueing discipline decreases mean waiting time by 5 or 30 minutes is uninteresting when studying coverage probabilities. We use the $\beta$ values of Kleijnen et al. (1985, p. 95); $\beta$ is displayed in appendix 4.

The users must also select $m_0$, the pilot sample size, in the two-stage and the sequential sampling procedures respectively. We study different values for $m_0$, namely 4, 9, and 25. These values seem realistic for experiments with complicated simulation models.

In summary, we have specified 27 ($= 3^3$) cases, which are defined by $\mathbf{X}$, $\mathbf{cov}(\mathbf{y})$, and $m_0$.

We repeat each case 150 times, since Kleijnen et al. (1985) found that 150 macro-replications 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 $2^{39}$. We employ different pseudorandom number streams for all 27 cases (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 Intermediate Monte Carlo Output on Underlying Assumptions

Before we present the final output, we discuss intermediate results that we obtained by running the simplest case 150 times using independent pseudorandom numbers (this case is only one of the 27 cases). This subsection may be skipped by readers interested in final outputs (estimated
coverage probabilities of the two sampling procedures), not in underlying assumptions (normality of average response, bias of variance estimator, correlations among averages, t distribution of Studentized β estimators).

First we test normality: does \( \bar{Y}_i(\bar{m}) \), the average response of combination \( i \) at the end of the two-stage procedure, have a Gaussian distribution? We have 150 observations on \( \bar{Y}_i(\bar{m}) \), which we use to form a histogram. We specify 14 classes for this histogram. We know the mean response (for example, the mean of the first combination is \( \beta_1 = \beta_2 = \beta_3 = -1 \)). However, we must estimate the variance. So to test goodness of fit we use a chi-square statistic with 12 (= 14 - 2) degrees of freedom. We select a type I error rate of 0.05, and test whether the histogram fits a normal distribution. For none of the \( n = 4 \) combinations we reject the null-hypothesis. Normal averages are also found for sequential sampling.

Our Monte Carlo experiment gives unbiased estimators of the (co)variances of the sample averages at the end of the two-stage procedure:

\[
\text{cov}(\bar{Y}_i(\bar{m}), \bar{Y}_j(\bar{m})) = \sum_{r=1}^{150} \{ \bar{Y}_u(\bar{m}) - \bar{Y}_i(\bar{m}) \} \{ \bar{Y}_p(\bar{m}) - \bar{Y}_j(\bar{m}) \} / 149
\]

with

\[
\bar{Y}_i(\bar{m}) = \sum_{r=1}^{150} \bar{Y}_u(\bar{m}) / 150.
\]

Simulation practitioners can not compute the estimators in (5.2), since they execute their simulation experiments only once (not 150 times). They do compute \( \text{var}(\bar{Y}_i(\bar{m})) \), the estimated variance of \( \bar{Y}_i(\bar{m}) \), which is biased (see the discussion of eq. 3.7). The example in Appendix 2 gives the following biased variance estimates, 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 respectively. To test whether these differences are significant, we can not apply a standard statistical procedure, We proceed as follows.

We compute \( \text{var}(\bar{Y}_i(\bar{m})) \), the average value of the 150 realizations of the simulationist’s estimator (3.7):

\[
\text{var}(\bar{Y}_i(\bar{m})) = \sum_{r=1}^{150} \text{var}(\bar{Y}_i(\bar{m})) / 150.
\]

We also compute \( \text{cov}(\bar{Y}_i(\bar{m}), \bar{Y}_j(\bar{m})) \), the one value obtained after our 150 Monte Carlo
experiments, which follows from (5.2) upon replacing \( j \) by \( i \). (We use the symbol \( \text{cov}(\bar{y}(\bar{m}), \bar{y}(\bar{m})) \) to avoid confusion with \( \text{var}(\bar{y}(\bar{m})) \) defined in equation 3.7.) It is well known that the variance of the estimated variance computed from \( n \) observations on a Gaussian variable with variance (say) \( \sigma^2 \) is \( 2\sigma^4/(n-1) \). Therefore we compute the statistic

\[
Z = \frac{\{\text{var}(\bar{y}(\bar{m})) - \text{cov}(\bar{y}(\bar{m}), \bar{y}(\bar{m}))\}}{\sqrt{\sum_{i=1}^{150} \{\text{var}(\bar{y}(\bar{m})) - \text{var}(\bar{y}(\bar{m}))\}^2/(149 \times 150) + 2 \text{cov}(\bar{y}(\bar{m}), \bar{y}(\bar{m}))^2/150}}.
\] (5.5)

Note that the last factor in (5.5) is 150: although \( \text{E}(\sigma^2) = \sigma^2 \), we know that \( \text{E}(\sigma^4) \neq \sigma^4 \) so the exact value of this factor is rather arbitrary.

We obtain the following \( n = 4 \) values for the test statistic in (5.5): 0.502, -1.631, -1.267, and -0.188. Assuming that this statistic is indeed standard normal, none of these four values is significant at a type I error rate of 0.10 (so certainly not at 0.05 or 0.01). So these results suggest that the (biased) variance estimator (3.7) has no important bias.

For the sequential procedure the users employ (4.6) instead of (3.7). It is straightforward to define a test statistic analogous to (5.5). This statistic gives the values: -0.675, -2.313, -1.067, and -0.545. Only one value is significant at a type I error rate of 0.10 (or 0.05). This one case does not suggest that the variance estimator has important bias. All four values of the statistic are negative, which confirms that this variance estimator has negative bias.

The simulationists can not estimate the correlations among the sample averages, from a single simulation experiment (there might be non-zero correlations, even if the simulationists employ independent seeds; see the discussion of equation 3.10). Therefore we assumed these correlations to be zero. This assumption is now investigated. 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 the regression parameters \( \hat{\beta} \) 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 test whether the Studentized \( \hat{\beta}_q \) has a \( t \) distribution. In the discussion leading to (3.11) we proposed degrees of freedom equal to \( m_0 - 1 \). So now we determine the histogram of \( t_{24} \). We again use 14 classes, which leads to a chi-square statistic with 13 degrees of freedom. For none of the three regression parameters we find a significant \( \chi^2 \). For the sequential procedure the same conclusion holds: the Studentized \( \hat{\beta}_q \) is distributed like \( t_{m_0 - 1} \). Because
the degrees of freedom are so high ($\nu = 24$), we might also conclude that the Studentized $\hat{\beta}_q$ are standard normal.

Simulation practitioners, however, use only the tails of the $t$ distribution; see (3.11). So as a whole the distribution of the Studentized $\beta$ 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.

5.3 Final Monte Carlo Output: Estimated Coverage Probabilities

We consider the two-sided $1 - \alpha$ confidence interval for $\beta_q$ based on the tabulated $1 - \alpha/2$ quantile (say) $t_{\alpha/2}$ of the Student statistic $t$ with $v = m_0 - 1$ or $v = \infty$; see (3.11). We suppose that analysts use one of the following classic values for $\alpha$: 0.01, 0.05 or 0.10. The analysts can apply a two-stage or a sequential sampling procedure. Whatever procedure they select, they must also select $m_0$, the pilot sample size. We conjecture that the statistical performance may depend on the variance heterogeneity (quantified by $\text{cov}(\gamma)$) and the metamodel specification and experimental design (defined by $X$). Therefore we study 27 cases, defined by $X$, $\text{cov}(\gamma)$, and $m_0$.

Each of the 150 macro-replications of a case yields one confidence interval per regression parameter $\beta_q$, given the values for $\alpha$ and $v$ and a sampling procedure. If that interval does not cover the true parameter value $\beta_q$, then we score (say) a one; otherwise a zero. This scoring defines the binomial variable (say) $\hat{p}_q$.

We formulate a one-sided null-hypothesis:

$$v_q: H_0: E(\hat{p}_q) \leq \alpha \text{ versus } H_1: E(\hat{p}_q) > \alpha \quad (5.6)$$

The alternative hypothesis means that the coverage probability is smaller than the nominal value $1 - \alpha$, that is, the confidence interval (3.11) is too tight.

We also test the related two-sided hypothesis

$$v_q: H'_0: E(\hat{p}_q) = \alpha \text{ versus } H'_1: E(\hat{p}_q) \neq \alpha \quad (5.7)$$

Because we obtained similar conclusions, we report only on the one-sided null-hypothesis, unless (5.6) and (5.7) give different conclusions.

To test $H_0$ it is convenient to approximate the binomial distribution of $\hat{p}_q$ through the
normal distribution $N(\alpha, \alpha(1-\alpha)/150)$ with $\alpha$ defined by (5.6).

A complication is that there are several parameters in the metamodel (1.1): $\beta_q$ with $q = 1, \ldots, Q$ (Q equals 13, 4, and 3 in our Monte Carlo experiment; see §5.1). We assume that the users want a confidence interval for each individual regression parameter that covers the true value with probability $1 - \alpha$; that is, the per comparison error rate has value $\alpha$. However, some thought shows that the more parameters there are in the metamodel (higher Q), the more likely it is that some estimated coverage probability significantly deviates from its expected value, by pure chance! We wish to keep the probability of rejecting $H_0$ in (5.6) below 0.05. Therefore we apply Bonferroni's inequality, that is, we test the coverage probability per individual parameter $\beta_q$ 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). So the more parameters there are in the metamodel, the larger the deviation of an estimated individual coverage probability from $1 - \alpha$ we accept.

In practice the users must select a sampling procedure (two-stage or sequential), a pilot sample size $m_0$, and the appropriate degrees of freedom $\nu$ ($m_0 - 1$ or $\infty$). For $\alpha$ they may either use a value that is traditional in their organization or they may try different values to see how these values affect their conclusions about the importance of the simulation parameters. To make our Monte Carlo results more general, we present results for different $\alpha$ values. The users' metamodel and experimental design are supposed to be given (see X), and so is $\text{cov}(\mathbf{y})$.

First we discuss the two-stage procedure. If the users base the confidence interval for $\beta_q$ on the $t$ distribution with $m_0 - 1$ degrees of freedom, then we reject $H_0$ in none of the 27 cases, except for 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).

If the users derive the confidence interval for $\beta_q$ from the normal distribution ($t_m$), then they get tighter intervals. Then for small pilot samples ($m_0$ is 4 or 9) the estimated coverage probabilities are significantly smaller than their nominal values, whatever $\alpha$ value the simulationists use (0.01, 0.05, or 0.10). For a 'large' pilot sample ($m_0 = 25$) the $t$ statistic with $m_0 - 1$ degrees of freedom approaches the standard normal distribution; then the estimated coverage probabilities are often not significantly small.

For the sequential procedure we get similar results. In other words, confidence intervals based on the normal distribution give too tight confidence intervals, except for cases with large pilot-samples. The $t$ distribution with $m_0 - 1$ degrees of freedom gives confidence intervals with estimated coverage probabilities that do not significantly deviate from the value $1 - \alpha$ specified by the users, except for one of the 27 cases (in that case an $\alpha$ of 0.05 or 0.10 gives significantly
low coverage probability; the less tight interval using an $\alpha$ of 0.01 does not).

In summary, both the two-stage procedure and the sequential procedure give 'acceptable' coverage probabilities, that is, confidence intervals for the regression parameters $\beta_q$ that do not deviate significantly from the user specified value $1 - \alpha$.

So now we compare the statistical efficiency of the two procedures. We measure this efficiency through the expected value of $m$ for the two-stage procedure and $m_T$ for the sequential procedure ($m$ denotes the number of simulation responses). To reduce sampling error we repeat all 27 cases 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 per combination of simulation inputs, in regression metamodeling. Both procedures determine the number of responses such that the variances of the average responses become 'approximately' constant. The first procedure is a two-stage rule; the second one is sequential.

After sampling (simulation) stops, Weighted Least Squares (WLS) is applied to the average simulation responses to estimate the metamodel parameters $\beta_q$ with $q = 1, \ldots, Q$. The WLS estimator is proved to be unbiased. A large Monte Carlo experiment shows that the $1 - \alpha$ confidence interval for $\beta_q$ based on the standard normal variable is too tight, that is, the estimated coverage probability significantly deviates from the prespecified $1 - \alpha$ value. However, this interval has 'acceptable' coverage probability (no significant deviation from $1 - \alpha$), when based on 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 procedures (two-stage and sequential).

The two-stage procedure is easier to understand and easier to program and implement. The sequential rule, however, requires fewer simulation responses.

In future research the robustness of the two rules with respect to nonnormality may be investigated. Further, MacNair and Welch (1991, p. 827) investigate sequential procedures that include testing the lack of fit of the metamodel. More research is needed for sample-size selection in metamodels for steady-state simulations analyzed by renewal analysis and other time-series analyses. In general, many statistical properties that are known for fixed sample sizes, must be re-evaluated for random sample sizes.

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 symbol $s_i^2 = s_i^2(m_o)$:

$$
\sum_{i=1}^{n} (y_i - \bar{y}_i(m_i))^2 = \sum_{i=1}^{n} (y_i - \bar{y}_i(m_o) + \bar{y}_i(m_o) - \bar{y}_i(m_i))^2 + \sum_{i=m_o+1}^{n} (y_i - \bar{y}_i(\ast)) + \bar{y}_i(\ast) - \bar{y}_i(m_i))^2
$$

$$
= \sum_{i=1}^{n} (y_i - \bar{y}_i(m_o))^2 + m_o(\bar{y}_i(m_o) - \frac{m_o}{m_i} \bar{y}_i(m_o) - \frac{m_i}{m_o} \bar{y}_i(\ast))^2
$$

$$
+ \sum_{i=m_o+1}^{n} (y_i - \bar{y}_i(\ast))^2 + (m_i - m_o)(\bar{y}_i(\ast) - \frac{m_i}{m_o} \bar{y}_i(m_i) - \frac{m_i}{m_o} \bar{y}_i(\ast))^2
$$

$$
= (m_o - 1) s_i^2 + \sum_{i=m_o+1}^{n} (y_i - \bar{y}_i(\ast))^2 + \frac{m_o(m_i - m_o)}{m_i} (\bar{y}_i(m_o) - \bar{y}_i(\ast))^2.
$$

Assume all $s_i^2$ with $i = 1, \ldots, n$ are fixed, so the $m_i$ are fixed. Then the expectation of $\sum_{i=1}^{n} (y_i - \bar{y}_i(m_i))^2$ conditioned on $s_i^2(m_o) = s_i^2(m_o)$ is

$$
E\{\sum_{i=1}^{n} (y_i - \bar{y}_i(m_i))^2 | s_i^2(m_o) = s_i^2(m_o), i = 1, \ldots, n\} = (m_o - 1) s_i^2 + \sum_{i=m_o+1}^{n} (\bar{y}_i(\ast))^2
$$

$$
+ \frac{m_o(m_i - m_o)}{m_i} (\frac{1}{m_i} s_i^2 + \frac{1}{m_i - m_o} \sigma_i^2)
$$

$$
= \frac{m_o(m_i - 1)}{m_i} s_i^2 + \frac{(m_i - m_o)(m_i - 1)}{m_i} \sigma_i^2.
$$

So we get

$$
E(s_i^2(m_i)) = \frac{m_o}{m_i} s_i^2 + \frac{(m_i - m_o)}{m_i} \sigma_i^2
$$

$$
= s_i^2 + \frac{(m_i - m_o)}{m_i} (\sigma_i^2 - s_i^2) \ \text{QED.}
$$

Appendix 2: Example of two-stage sampling

We consider an example with two simulation inputs (in a more realistic simulation these two inputs might be traffic rate and queueing priority). Suppose that all inputs have effects of +1; the overall mean or intercept is also +1 (so we select $\beta_1 = \beta_2 = \beta_3 = 1$). Suppose the users employ a $2^2$ design to specify the four input combinations of their 'simulation'. Hence $X$ in the metamodel (1.1) has column 1 with all elements equal to 1; column 2 with elements -1, 1, -1, 1; and
column 3 with -1, -1, 1, and 1. (Consequently the expected value of the first combination is \( E(y) = x' \beta = \beta_1 - \beta_2 - \beta_3 = -1 \).)

For the response variances we select the values 1, 4, 8, and 11.38 respectively (\( \sigma_1^2 = 1, \ldots, \sigma_4^2 = 11.38 \)). For the pilot sample-size we take \( m_0 = 25 \).

To save space we do not display the \( m_0 \) individual responses for the four combinations; we mention only that the responses in combination 1 range between -2.25 and +1.32 (remember their mean is \( E(y) = -1 \) and their variance is 1). Their average is \( \bar{y} = -0.546 \) and their estimated sample variance is \( s_1^2(m_0) = 0.803 \). For the other three combinations we get the following variance estimators: \( s_2^2(m_0) = 4.915, s_3^2(m_0) = 8.816, \) and \( s_4^2(m_0) = 18.615 \) (the true variances are 4, 8, and 11.38 respectively).

These stage 1 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 'approximately equal': 0.03213, 0.02492, 0.02880, and 0.02283.

The estimated regression parameters are 1.056, 0.831, and 0.879 (their true values are 1). The standard errors are 0.08 for all three parameters.

Finally we consider the confidence intervals. We concentrate on the tightest interval, based on the normal approximation \( (u = \infty) \) and a high type I error probability \( \alpha = 0.10 \). Then \( z_{10} = 1.645 \) and the upper bound for \( \beta_3 \) (the parameter with the largest deviation from its mean) is \( 0.831 + (1.645)(0.08) \) is 0.96. This interval does not cover the true value, +1. Is this bad luck? This question is answered in §5.

Appendix 3: Example of sequential sampling

We consider the same example as in two-stage sampling (Appendix 2), with a new random number seed, which yields Table 1. So the pilot sample is 25. This yields \( s_1^2(m_0) = 1.165, s_2^2(m_0) = 2.881, s_3^2(m_0) = 7.495, \) and \( s_4^2(m_0) = 5.849 \). These estimates lead to one new observation from combinations 2, 3, and 4 respectively. This yields \( \bar{y}(\hat{m}_1 = 25) = -0.782, \bar{y}(\hat{m}_2 = 26) = 0.945, \bar{y}(\hat{m}_3 = 26) = 1.661, \bar{y}(\hat{m}_4 = 26) = 3.147 \). The re-estimated variances of the individual responses are 1.165, 2.967, 7.228, and 6.201. These estimates imply estimated required sample sizes of 25, 64, 155, and 133. So the procedure takes one more observation from combinations 2, 3, and 4. And so on.

After 221 stages the available number of observations is at least as big as the required number, for each combination. 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.

We observe that at the end of stage #220 the procedure estimates that the number of required responses for combination 4 is \((11.446/1.165)25 = 246\). The number 246 is entered under stage 220; at the end of stage #221 the procedure computes \( \hat{m}_4 = [(11.405/1.165)25] = 245 \) and stops.
At the end the mean response for combination 2 is estimated from 89 (not 90) responses. In stage #65 this number of responses was reached; then the average was 1.014 (not 1.012; see stage #66). The variance is estimated from all 90 observations: it is 4.156.

At the end 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 estimated regression parameters are 0.964, 0.883, and 0.848 with standard errors 0.11. The tightest confidence interval does cover the true value.

Let us compare these results with those of the two-stage procedure. The total number of simulation responses is 1031 in two-stage sampling and 503 in sequential sampling. Consequently the standard errors of the estimated regression parameters are larger in sequential sampling: 0.11 versus 0.08. A criterion for comparing the two procedures may be

\[ \sqrt{\text{var}(\hat{b})} \times \sum_{i=1}^{n} m_{it} \]

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 is answered in §5.

Appendix 4: 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 \( \beta \) in Case 1 from the Rotterdam harbor study in Kleijnen, Van den Burg and Van der Ham (1979).

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

\[ H = 0: \sigma_i^2 = 1 \text{ for all } i. \]
\[ H = 4: \sigma_i^2 = (4, 4.5, 5, 6, 8, 9, 10, 11, 12, 13.5, 14, 14.5, 16, 18, 19.5, 20). \]
\[ H = 11.84: \sigma_i^2 = (1, 2, 3, 4, 4.5, 5, 6, 7, 7.5, 8, 9, 9.5, 10, 11, 12, 12.84). \]

Case 2: \( \beta^* = (-1.42, -0.769, 13.44, -11.508). \)

\[ H = 0: \sigma_i^2 = 1 \text{ for all } i. \]
\[ H = 4: \sigma_i^2 = (4, 6, 8, 10, 12, 14, 16, 20). \]
\[ H = 10.83: \sigma_i^2 = (1, 2, 4, 5, 6, 7, 9, 11.83). \]

Case 3: \( \beta^* = (1, 1, 1). \)

\[ H = 0: \sigma_i^2 = 1 \text{ for all } i. \]
\[ H = 4: \sigma_i^2 = (4, 10, 16, 20). \]
\[ H = 10.38; \sigma_i^2 = (1, 4, 8, 11.38). \]

**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**

### TABLE I

**Example Data for Sequential Sampling**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Comb.</th>
<th>Available</th>
<th>Required</th>
<th>Mean $\bar{y}$</th>
<th>Variance $s^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>25</td>
<td>25</td>
<td>-0.782</td>
<td>1.165</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>26</td>
<td>64</td>
<td>0.945</td>
<td>2.967</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>26</td>
<td>155</td>
<td>1.661</td>
<td>7.228</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>26</td>
<td>133</td>
<td>3.147</td>
<td>6.201</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>25</td>
<td>25</td>
<td>-0.782</td>
<td>1.165</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>27</td>
<td>66</td>
<td>0.857</td>
<td>3.066</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>27</td>
<td>153</td>
<td>1.579</td>
<td>7.132</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>27</td>
<td>144</td>
<td>2.980</td>
<td>6.710</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>65</td>
<td>1</td>
<td>25</td>
<td>25</td>
<td>-0.782</td>
<td>1.165</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>89</td>
<td>90</td>
<td>1.014</td>
<td>4.203</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>89</td>
<td>126</td>
<td>1.005</td>
<td>5.892</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>89</td>
<td>180</td>
<td>2.927</td>
<td>8.380</td>
</tr>
<tr>
<td>66</td>
<td>1</td>
<td>25</td>
<td>25</td>
<td>-0.782</td>
<td>1.165</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>90</td>
<td>89</td>
<td>1.012</td>
<td>4.156</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>90</td>
<td>125</td>
<td>1.007</td>
<td>5.826</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>90</td>
<td>178</td>
<td>2.912</td>
<td>8.304</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>220</td>
<td>1</td>
<td>25</td>
<td>25</td>
<td>-0.782</td>
<td>1.165</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>90</td>
<td>89</td>
<td>1.012</td>
<td>4.156</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>143</td>
<td>142</td>
<td>0.945</td>
<td>6.623</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>244</td>
<td>246</td>
<td>2.684</td>
<td>11.446</td>
</tr>
<tr>
<td>221</td>
<td>1</td>
<td>25</td>
<td>25</td>
<td>-0.782</td>
<td>1.165</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>90</td>
<td>89</td>
<td>1.012</td>
<td>4.156</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>143</td>
<td>142</td>
<td>0.945</td>
<td>6.623</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>245</td>
<td>245</td>
<td>2.679</td>
<td>11.405</td>
</tr>
</tbody>
</table>
Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

<table>
<thead>
<tr>
<th>No.</th>
<th>Author(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>9324</td>
<td>H. Huizinga</td>
<td>The Financing and Taxation of U.S. Direct Investment Abroad</td>
</tr>
<tr>
<td>9326</td>
<td>T.C. To</td>
<td>Infant Industry Protection with Learning-by-Doing</td>
</tr>
<tr>
<td>9327</td>
<td>J.P.J.F. Scheepens</td>
<td>Bankruptcy Litigation and Optimal Debt Contracts</td>
</tr>
<tr>
<td>9328</td>
<td>T.C. To</td>
<td>Tariffs, Rent Extraction and Manipulation of Competition</td>
</tr>
<tr>
<td>9329</td>
<td>F. de Jong, T. Nijman and A. Röell</td>
<td>A Comparison of the Cost of Trading French Shares on the Paris Bourse and on SEAQ International</td>
</tr>
<tr>
<td>9330</td>
<td>H. Huizinga</td>
<td>The Welfare Effects of Individual Retirement Accounts</td>
</tr>
<tr>
<td>9331</td>
<td>H. Huizinga</td>
<td>Time Preference and International Tax Competition</td>
</tr>
<tr>
<td>9332</td>
<td>V. Feltkamp, A. Koster, A. van den Nouweland, P. Borm and S. Tijs</td>
<td>Linear Production with Transport of Products, Resources and Technology</td>
</tr>
<tr>
<td>9333</td>
<td>B. Lauterbach and U. Ben-Zion</td>
<td>Panic Behavior and the Performance of Circuit Breakers: Empirical Evidence</td>
</tr>
<tr>
<td>9334</td>
<td>B. Melenberg and A. van Soest</td>
<td>Semi-parametric Estimation of the Sample Selection Model</td>
</tr>
<tr>
<td>9335</td>
<td>A.L. Bovenberg and F. van der Ploeg</td>
<td>Green Policies and Public Finance in a Small Open Economy</td>
</tr>
<tr>
<td>9336</td>
<td>E. Schaling</td>
<td>On the Economic Independence of the Central Bank and the Persistence of Inflation</td>
</tr>
<tr>
<td>9337</td>
<td>G.-J. Otten</td>
<td>Characterizations of a Game Theoretical Cost Allocation Method</td>
</tr>
<tr>
<td>9338</td>
<td>M. Gradstein</td>
<td>Provision of Public Goods With Incomplete Information: Decentralization vs. Central Planning</td>
</tr>
<tr>
<td>9339</td>
<td>W. Güth and H. Kliemt</td>
<td>Competition or Co-operation</td>
</tr>
<tr>
<td>9340</td>
<td>T.C. To</td>
<td>Export Subsidies and Oligopoly with Switching Costs</td>
</tr>
<tr>
<td>9341</td>
<td>A. Demirgüç-Kunt and H. Huizinga</td>
<td>Barriers to Portfolio Investments in Emerging Stock Markets</td>
</tr>
<tr>
<td>9342</td>
<td>G.J. Almekinders</td>
<td>Theories on the Scope for Foreign Exchange Market Intervention</td>
</tr>
<tr>
<td>No.</td>
<td>Author(s)</td>
<td>Title</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>9343</td>
<td>E.R. van Dam and W.H. Haemers</td>
<td>Eigenvalues and the Diameter of Graphs</td>
</tr>
<tr>
<td>9344</td>
<td>H. Carlsson and S. Dasgupta</td>
<td>Noise-Proof Equilibria in Signaling Games</td>
</tr>
<tr>
<td>9345</td>
<td>F. van der Ploeg and A.L. Bovenberg</td>
<td>Environmental Policy, Public Goods and the Marginal Cost of Public Funds</td>
</tr>
<tr>
<td>9346</td>
<td>J.P.C. Blanc and R.D. van der Mei</td>
<td>The Power-series Algorithm Applied to Polling Systems with a Dormant Server</td>
</tr>
<tr>
<td>9347</td>
<td>J.P.C. Blanc</td>
<td>Performance Analysis and Optimization with the Power-series Algorithm</td>
</tr>
<tr>
<td>9348</td>
<td>R.M.W.J. Beetsma and F. van der Ploeg</td>
<td>Intramarginal Interventions, Bands and the Pattern of EMS Exchange Rate Distributions</td>
</tr>
<tr>
<td>9349</td>
<td>A. Simonovits</td>
<td>Intercohort Heterogeneity and Optimal Social Insurance Systems</td>
</tr>
<tr>
<td>9350</td>
<td>R.C. Douven and J.C. Engwerda</td>
<td>Is There Room for Convergence in the E.C.?</td>
</tr>
<tr>
<td>9351</td>
<td>F. Vella and M. Verbeek</td>
<td>Estimating and Interpreting Models with Endogenous Treatment Effects: The Relationship Between Competing Estimators of the Union Impact on Wages</td>
</tr>
<tr>
<td>9352</td>
<td>C. Meghir and G. Weber</td>
<td>Intertemporal Non-separability or Borrowing Restrictions? A Disaggregate Analysis Using the US CEX Panel</td>
</tr>
<tr>
<td>9353</td>
<td>V. Feltkamp</td>
<td>Alternative Axiomatic Characterizations of the Shapley and Banzhaf Values</td>
</tr>
<tr>
<td>9355</td>
<td>Z. Yang</td>
<td>A Simplicial Algorithm for Computing Robust Stationary Points of a Continuous Function on the Unit Simplex</td>
</tr>
<tr>
<td>9356</td>
<td>E. van Damme and S. Hurkens</td>
<td>Commitment Robust Equilibria and Endogenous Timing</td>
</tr>
<tr>
<td>9357</td>
<td>W. Güth and B. Peleg</td>
<td>On Ring Formation In Auctions</td>
</tr>
<tr>
<td>9358</td>
<td>V. Bhaskar</td>
<td>Neutral Stability In Asymmetric Evolutionary Games</td>
</tr>
<tr>
<td>9359</td>
<td>F. Vella and M. Verbeek</td>
<td>Estimating and Testing Simultaneous Equation Panel Data Models with Censored Endogenous Variables</td>
</tr>
<tr>
<td>9360</td>
<td>W.B. van den Hout and J.P.C. Blanc</td>
<td>The Power-Series Algorithm Extended to the BMAP/PH/1 Queue</td>
</tr>
<tr>
<td>9361</td>
<td>R. Heuts and J. de Klein</td>
<td>An (s,q) Inventory Model with Stochastic and Interrelated Lead Times</td>
</tr>
<tr>
<td>No.</td>
<td>Author(s)</td>
<td>Title</td>
</tr>
<tr>
<td>------</td>
<td>------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>9362</td>
<td>K.-E. Wärneryd</td>
<td>A Closer Look at Economic Psychology</td>
</tr>
<tr>
<td>9363</td>
<td>P.J.-J. Herings</td>
<td>On the Connectedness of the Set of Constrained Equilibria</td>
</tr>
<tr>
<td>9364</td>
<td>P.J.-J. Herings</td>
<td>A Note on &quot;Macroeconomic Policy in a Two-Party System as a Repeated Game&quot;</td>
</tr>
<tr>
<td>9365</td>
<td>F. van der Ploeg and A. L. Bovenberg</td>
<td>Direct Crowding Out, Optimal Taxation and Pollution Abatement</td>
</tr>
<tr>
<td>9366</td>
<td>M. Pradhan</td>
<td>Sector Participation in Labour Supply Models: Preferences or Rationing?</td>
</tr>
<tr>
<td>9367</td>
<td>H.G. Bloemen and A. Kapteyn</td>
<td>The Estimation of Utility Consistent Labor Supply Models by</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Means of Simulated Scores</td>
</tr>
<tr>
<td>9368</td>
<td>M.R. Baye, D. Kovenock and C.G. de Vries</td>
<td>The Solution to the Tullock Rent-Seeking Game When R &gt; 2: Mixed-Strategy Equilibria and Mean Dissipation Rates</td>
</tr>
<tr>
<td>9369</td>
<td>T. van de Klundert and S. Smulders</td>
<td>The Welfare Consequences of Different Regimes of Oligopolistic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Competition in a Growing Economy with Firm-Specific Knowledge</td>
</tr>
<tr>
<td>9370</td>
<td>G. van der Laan and D. Talman</td>
<td>Intersection Theorems on the Simplotope</td>
</tr>
<tr>
<td>9371</td>
<td>S. Muto</td>
<td>Alternating-Move Preplays and vN - M Stable Sets in Two Person Strategic Form Games</td>
</tr>
<tr>
<td>9372</td>
<td>S. Muto</td>
<td>Voters' Power in Indirect Voting Systems with Political Parties:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>the Square Root Effect</td>
</tr>
<tr>
<td>9373</td>
<td>S. Smulders and R. Gradus</td>
<td>Pollution Abatement and Long-term Growth</td>
</tr>
<tr>
<td>9374</td>
<td>C. Fernandez, J. Osiewalski and M.F.J. Steel</td>
<td>Marginal Equivalence in v-Spherical Models</td>
</tr>
<tr>
<td>9375</td>
<td>E. van Damme</td>
<td>Evolutionary Game Theory</td>
</tr>
<tr>
<td>9376</td>
<td>P.M. Kort</td>
<td>Pollution Control and the Dynamics of the Firm: the Effects of Market Based Instruments on Optimal Firm Investments</td>
</tr>
<tr>
<td>9377</td>
<td>A. L. Bovenberg and F. van der Ploeg</td>
<td>Optimal Taxation, Public Goods and Environmental Policy with</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Involuntary Unemployment</td>
</tr>
<tr>
<td>9378</td>
<td>F. Thuijsman, B. Peleg, M. Amitai &amp; A. Shmida</td>
<td>Automata, Matching and Foraging Behavior of Bees</td>
</tr>
<tr>
<td>9379</td>
<td>A. Lejour and H. Verbon</td>
<td>Capital Mobility and Social Insurance in an Integrated Market</td>
</tr>
<tr>
<td>No.</td>
<td>Author(s)</td>
<td>Title</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>9380</td>
<td>C. Fernandez, J. Osiewalski and M. Steel</td>
<td>The Continuous Multivariate Location-Scale Model Revisited: A Tale of Robustness</td>
</tr>
<tr>
<td>9381</td>
<td>F. de Jong</td>
<td>Specification, Solution and Estimation of a Discrete Time Target Zone Model of EMS Exchange Rates</td>
</tr>
<tr>
<td>9401</td>
<td>J.P.C. Kleijn and R.Y. Rubinstein</td>
<td>Monte Carlo Sampling and Variance Reduction Techniques</td>
</tr>
<tr>
<td>9402</td>
<td>F.C. Drost and B.J.M. Werker</td>
<td>Closing the Garch Gap: Continuous Time Garch Modeling</td>
</tr>
<tr>
<td>9403</td>
<td>A. Kapteyn</td>
<td>The Measurement of Household Cost Functions: Revealed Preference Versus Subjective Measures</td>
</tr>
<tr>
<td>9404</td>
<td>H.G. Bloemen</td>
<td>Job Search, Search Intensity and Labour Market Transitions: An Empirical Exercise</td>
</tr>
<tr>
<td>9405</td>
<td>P.W.J. De Bijl</td>
<td>Moral Hazard and Noisy Information Disclosure</td>
</tr>
<tr>
<td>9406</td>
<td>A. De Waegenaere</td>
<td>Redistribution of Risk Through Incomplete Markets with Trading Constraints</td>
</tr>
<tr>
<td>9407</td>
<td>A. van den Nouweland, P. Borm, W. van Golstein Brouwers, R. Groot Bruinderink, and S. Tijs</td>
<td>A Game Theoretic Approach to Problems in Telecommunication</td>
</tr>
<tr>
<td>9408</td>
<td>A.L. Bovenberg and F. van der Ploeg</td>
<td>Consequences of Environmental Tax Reform for Involuntary Unemployment and Welfare</td>
</tr>
<tr>
<td>9409</td>
<td>P. Smit</td>
<td>Arnoldi Type Methods for Eigenvalue Calculation: Theory and Experiments</td>
</tr>
<tr>
<td>9410</td>
<td>J. Eichberger and D. Kelsey</td>
<td>Non-additive Beliefs and Game Theory</td>
</tr>
<tr>
<td>9411</td>
<td>N. Dagan, R. Serrano and O. Volij</td>
<td>A Non-cooperative View of Consistent Bankruptcy Rules</td>
</tr>
<tr>
<td>9412</td>
<td>H. Bester and E. Petrakis</td>
<td>Coupons and Oligopolistic Price Discrimination</td>
</tr>
<tr>
<td>9413</td>
<td>G. Koop, J. Osiewalski and M.F.J. Steel</td>
<td>Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function</td>
</tr>
<tr>
<td>9414</td>
<td>C. Kilby</td>
<td>World Bank-Borrower Relations and Project Supervision</td>
</tr>
<tr>
<td>9415</td>
<td>H. Bester</td>
<td>A Bargaining Model of Financial Intermediation</td>
</tr>
<tr>
<td>No.</td>
<td>Author(s)</td>
<td>Title</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>9417</td>
<td>J. de la Horra and C. Fernandez</td>
<td>Sensitivity to Prior Independence via Farlie-Gumbel-Morgenstern Model</td>
</tr>
<tr>
<td>9418</td>
<td>D. Tolman and Z. Yang</td>
<td>A Simplicial Algorithm for Computing Proper Nash Equilibria of Finite Games</td>
</tr>
<tr>
<td>9419</td>
<td>H.J. Bierens</td>
<td>Nonparametric Cointegration Tests</td>
</tr>
<tr>
<td>9420</td>
<td>G. van der Laan, D. Talman and Z. Yang</td>
<td>Intersection Theorems on Polytopes</td>
</tr>
<tr>
<td>9421</td>
<td>R. van den Brink and R.P. Gilles</td>
<td>Ranking the Nodes in Directed and Weighted Directed Graphs</td>
</tr>
<tr>
<td>9422</td>
<td>A. van Soest</td>
<td>Youth Minimum Wage Rates: The Dutch Experience</td>
</tr>
<tr>
<td>9423</td>
<td>N. Dagan and O. Volij</td>
<td>Bilateral Comparisons and Consistent Fair Division Rules in the Context of Bankruptcy Problems</td>
</tr>
<tr>
<td>9424</td>
<td>R. van den Brink and P. Borm</td>
<td>Digraph Competitions and Cooperative Games</td>
</tr>
<tr>
<td>9425</td>
<td>P.H.M. Ruys and R.P. Gilles</td>
<td>The Interdependence between Production and Allocation</td>
</tr>
<tr>
<td>9426</td>
<td>T. Callan and A. van Soest</td>
<td>Family Labour Supply and Taxes in Ireland</td>
</tr>
<tr>
<td>9427</td>
<td>R.M.W.J. Beetsma and F. van der Ploeg</td>
<td>Macroeconomic Stabilisation and Intervention Policy under an Exchange Rate Band</td>
</tr>
<tr>
<td>9428</td>
<td>J.P.C. Kleijnen and W. van Groenendaal</td>
<td>Two-stage versus Sequential Sample-size Determination in Regression Analysis of Simulation Experiments</td>
</tr>
</tbody>
</table>