Proceedings of the 1998 Winter Simulation Conference
D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.

BOOTSTRAPPING AND VALIDATION OF METAMODELS IN SIMULATION

Jack P.C. Kleijnen
Ad J. Feelders

Department of Information Systems and Auditing
(BIKA)/Center for Economic Research (CentER)
Tilburg University (Katholieke Universiteit Brabant)
5000 LE Tilburg, NETHERLANDS

Russell C.H. Cheng

Business School
University of Kent at Canterbury
Kent CT2 7NF, Canterbury, U.K.

ABSTRACT

Bootstrapping is a resampling technique that requires less computer time than simulation does. Bootstrapping -like simulation- must be defined for each type of application. This paper defines bootstrapping for random simulations with replicated runs. The focus is on linear regression metamodels. The metamodel's parameters are estimated through Generalized Least Squares. Its fit is measured through Rao's lack-of-fit F-statistic. The distributions of this statistic is estimated through bootstrapping. The main conclusions are (i) not the regression residuals should be bootstrapped; instead the deviations that also occur in the standard deviation, should be bootstrapped (ii) bootstrapping Rao's lack-of-fit statistic is a good alternative to the F-test: it gives virtually identical results when the assumptions of the F-test are known to apply, and somewhat better results otherwise.

1 INTRODUCTION

In this paper we examine the validation of a linear regression model used as a *metamodel* or response surface (an approximation of the input/output or I/O transformation implied by the underlying simulation model).Different types of metamodels and their validation are surveyed in Kleijnen and Sargent (1997). We, however, introduce *bootstrapping* as a technique for this validation. Bootstrapping outside simulation is studied in the seminal book, Efron and Tibshirani (1993), which we abbreviate to E & T. In general, bootstrapping means that the data (say) $\boldsymbol{z}_j$ in the original sample of size s are randomly resampled with replacement ($j = 1, ..., s$). E & T (pp. 115, 383) comment that 'bootstrapping is not a uniquely defined concept ... alternative bootstrap methods may coexist'. We shall give our interpretation of bootstrapping for our problem.

We focus on Rao's lack-of-fit F-statistic, but we plan to study several other popular statistics, namely the coefficient of determination (R^2), possibly adjusted for the number of parameters (R^2_{adj}), the linear correlation coefficient, also known as Pearson's rho (ρ), and cross-validation giving Studentized prediction errors and using Bonferroni's inequality (resulting in, say, $\tilde{t}_{\max}$). These various statistics can be studied through a Monte Carlo experiment. This paper gives preliminary results (at the WSC '98 conference more extensive results will be presented; see §5 on future research).

The *literature* gives the following picture. Stine (1985) also examines bootstrapping in regression analysis, but he assumes that the regression model is correct; moreover he assumes constant response variances, whereas simulation applications show variance heterogeneity, in general. Breiman (1992) investigates the selection of the correct regression model, but he assumes no replications ($m_i = 1$), constant response variances, and a particular parametric bootstrap of a particular statistic different from ours. Bootstrapping has not yet been applied frequently in systems simulation; two academic examples are provided in Friedman and Friedman (1995), including references to software that permits bootstrapping (they mention SAS and SPSS; we use S-Plus). Kim, Willemain, Haddock, and Runger (1993) formulate their so-called 'threshold' bootstrap for the analysis of autocorrelated simulation outputs. Barton and Schruben (1996), Cheng and Holland (1997), and Cheng (1995) investigate bootstrapping of empirical input distributions in simulation.

The main conclusions are (i) not the regression residuals $\hat{e}$ should be bootstrapped; instead the deviations $w - \bar{w}$ should be bootstrapped (ii) bootstrapping Rao's lack-of-fit statistic is a good alternative to the F-test: it gives virtually identical results when the assumptions of the F-test are known to apply, and somewhat better results otherwise.

We *organize* the remainder of this paper as follows. In §2 we summarize linear regression metamodels with parameters estimated through Least Squares (LS); we define

Rao's lack-of-fit F-statistic. In §3 we discuss how the distribution of this statistic can be estimated through bootstrapping. In §4 we present preliminary Monte Carlo results. In §5 we briefly discuss future research. In §6 we summarize the main conclusions.

Note: Bootstrapping in simulation raises an interesting question. Instead of using the computer to generate responses through bootstrapping, the computer may be used to generate more simulation responses, either for old factor combinations or for new combinations. In practice, many simulation models require much more computer time than regression analysis does. In those situations it makes sense indeed to bootstrap. Breiman (1992, p. 750) also discusses bootstrapping versus replicating, but not in a simulation context.

2 LINEAR REGRESSION METAMODELS IN SIMULATION

We first define some symbols. We use Greek letters for parameters; bold face for matrixes and vectors. We suppose that the simulation model has k factors, denoted by d_j with $j = 1, ..., k$ so $\boldsymbol{d} = (d_1, d_2, ..., d_k)$. We assume a single type of simulation response, denoted by w (e.g. average waiting time). We let n denote the total number of factor combinations actually simulated. Factor combination i with $i = 1, ..., n$ is replicated m_i times using non-overlapping pseudo-random number streams. This yields simulation response $w_{i;r}$ with $r = 1, .., m_i$. Hence, $N = \sum_{i=1}^{n} m_i$ denotes the total number of simulation observations. Let $\bar{\boldsymbol{w}}$ denote the n-dimensional vector of average simulation outputs that can be obtained from $w_{i;r}$. Denote the total number of parameters in the regression metamodel by q. For example, a first-order polynomial regression metamodel for k factors has parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_k)'$ so $q = k + 1$. Let $\boldsymbol{X}$ denote the $n \times q$ matrix of simulated independent regression variables; for example, its first column consists of ones; its first row of $\mathbf{x}_1 = (1, x_{1;2}, ... x_{1;q-1})$. We further assume that the n combinations of simulation factors are selected though application of the statistical theory on Design Of Experiments (DOE); for example, a 2^{k-p} design may be used. Hence $\boldsymbol{X}$ is controlled by the $n \times k$ design matrix $\boldsymbol{D} = (\boldsymbol{d}_i)$ and the form of the metamodel assumed. Consequently, in our situation the original (non-bootstrapped) data $\boldsymbol{Z}$ (see §1) consist of $(\boldsymbol{X}, \boldsymbol{w})$. Finally, let $\boldsymbol{y}$ denote the output for the correct (adequate, valid) metamodel, so

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e} \tag{1}$$

where $\boldsymbol{e}$ denotes the additive random residual with zero mean and covariance matrix $\boldsymbol{cov}(\boldsymbol{y})$.

An example of such a metamodel is a second-order polynomial. This metamodel has (i) one grand or overall mean β_0, (ii) k main effects β_j, (iii) $k(k-1)/2$ two-factor interactions $\beta_{j;j'}$ ($j < j'$ and $j' = 2, ..., k$), and (iv) k quadratic effects, $\beta_{j;j}$. Metamodel validation often concerns the selection of the correct degree of the polynomial. Cheng and Kleijnen (1997) gives a Bayesian approach, whereas this paper gives a bootstrap approach. Kleijnen and Standridge (1987) applies a popular ad hoc approach.

First suppose that the metamodel is fitted applying *ordinary LS* or OLS. Then the OLS estimator (say) $\hat{\beta}$ for a first-order polynomial (meta)model is

$$\hat{\beta} = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\bar{\boldsymbol{w}}. \tag{2}$$

This formula assumes that the inverse exists; DOE ensures that $\boldsymbol{X}_1$ is indeed not collinear. For example, if $k = 3$ and a first-order polynomial is assumed, then a 2^{3-1} design gives an orthogonal $\boldsymbol{X}_1$, assuming the factors are standardized; see Kleijnen and Bettonvil (1990).

Generalized LS (GLS) estimates $\boldsymbol{\beta}$ accounting for variance heterogeneity (resulting in WLS) and -in case of common random numbers- correlations among simulation responses. These variances and covariances can be estimated by

$$s_{i;i'}(w) = \frac{\sum_{r=1}^{m} (w_{i;r} - \bar{w}_i)(w_{i';r} - \bar{w}_{i'})}{(m_i - 1)} \quad with\ i, i' = 1, ..., n. \tag{3}$$

Sometimes we shall abbreviate $s_{i;i'}(w)$ to $s_{i;i'}$; further $s_{i;i'} = s_{i';i}$ and $s_{i;i} \equiv s_i^2$; in case of common random numbers we have constant replication numbers, $m_i = m$. E & T (p. 53) use 'plug in' estimators with denominators m_i, but we prefer unbiased estimators.

Using the covariance estimators in (3), we define the $n \times n$ matrix $\boldsymbol{S}_w = (s_{i;i'})$. This random matrix is used in Estimated GLS or EGLS:

$$\tilde{\beta} \propto (\boldsymbol{X}'\boldsymbol{S}_w^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{S}_w^{-1}\bar{\boldsymbol{w}}. \tag{4}$$

Notice that this is a non-linear estimator since it uses the random variables $\boldsymbol{S}_w$ and $\bar{\boldsymbol{w}}$.

To explain Rao's test we start as follows. The literature on DOE often assumes replications, and the concomitant Analysis of Variance (ANOVA) assumes normally,

independently, and identically (NID) distributed residuals with zero mean: $\mathbf{e} \sim NID(0, \sigma^2)$; see (1). ANOVA then uses the classical *F lack-of-fit test*, which compares two estimators of the common response variance σ^2, as follows (i) The first estimator uses s_i^2 with $i = 1, ..., n$, which are the classic variance estimators based on replication defined in (3). Because the true response variance σ^2 is constant, these n estimators are averaged or pooled: $\overline{s^2} = \sum_{i=1}^{n} s_i^2/n$. (If m_i is not constant, then a weighted average is used, with the degrees of freedom m_i - 1 as weights.)
(ii) The second estimator uses the n estimated residuals where the linear regression model (1) together with OLS

$$\bar{e}_i = \bar{w}_i - \hat{y}_i \quad (i = 1, ..., n) \tag{5}$$

gives the predictor or forecast $\hat{y}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}}$. These residuals give the second variance estimator, $\sum_{i=1}^{n} \bar{e}_i^2 m/(n-q)$ This estimator is unbiased if and only if (iff) the regression model is specified correctly; otherwise this estimator overestimates the true variance.
(iii) Finally, these two estimators are compared statistically through the F-statistic $F_{n-q, N-n}$. The lack of fit is declared significant if this statistic exceeds its upper 1 - α quantile.

Rao (1959) extends this test from OLS to EGLS, assuming a constant number of correlated replications, as is the case for common random numbers. Let $\tilde{\mathbf{e}}$ denote the n-dimensional vector with estimated EGLS residuals. Then the test statistic becomes

$$F_{n-q, m-n+q} = \frac{\tilde{\mathbf{e}}'\mathbf{S}_w^{-1}\tilde{\mathbf{e}} \; m(m-n+q)}{[(n-q)(m-1)]}. \tag{6}$$

Kleijnen (1992) shows that this test performs well for symmetrically distributed simulation responses; for example, it works for normally or uniformly distributed, but not for lognormally distributed simulation responses. In queueing simulations the responses may indeed be asymmetrically distributed; also see Cheng and Kleijnen (1997). Fortunately, bootstrapping permits the estimation of any statistic including (6), for any distribution.

3 BOOTSTRAPPING RAO'S LACK-OF-FIT F-STATISTIC

E & T describe the real world by $\mathbf{z} \sim P$ where $\mathbf{z}$ is an independently and identically distributed (i.i.d.) variable (possibly, multi-variate), and P is its distribution function.(Hence, each individual observation $\mathbf{z}_j$ may be sampled 0, 1, ..., s times; so this sampling follows a multinomial probability function). Let $\hat{F}$ denote the estimated probability function. Then the bootstrap sample is

$$\mathbf{z}_j^* \sim \hat{F} \quad \text{where } \hat{F} = 1/s \quad (j = 1, ..., s). \tag{7}$$

Bootstrapping supposes that the data are summarized by a *statistic* (say) $\hat{\theta} = s(\mathbf{z})$. E & T gives as examples the mean, variance, linear correlation coefficient, and eigenvalues; we focus on Rao's statistic (6).

If no common random numbers are used, then we allow different replication numbers $m_i \neq m$. Such unequal replication numbers are used in, for example, Cheng and Kleijnen (1997, 1998), assuming that queueing systems with n different traffic rates are simulated with more customers when traffic rates are high.

Common random numbers create positive correlations between the components of $\mathbf{w}_r$ with $\mathbf{w}_r = (w_{1;r}, w_{2;r}, ..., w_{n;r})'$ and $r = 1, .., m$; that is, we assume a constant number of replications $m_i = m$.

Note: E & T (p. 111) considers regression analysis without DOE, so there are no replications: $m_i = 1$ so $N = n$. In our proposed bootstrap, a restriction is that the resulting matrix $\mathbf{X}$ be non-singular. This restriction, however, is satisfied as DOE guarantees a non-singular $n \times q$ matrix $\mathbf{X}$.

We present one bootstrap technique, namely a non-parametric version. We ensure that in our bootstrap the metamodel is correct, so that we obtain a bootstrap distribution of Rao's statistic under the null-hypothesis of a valid metamodel. With the resulting bootstrap distribution we can confront the original statistic F computed for the metamodel from the original I/O simulation data ($\mathbf{X}$, $\mathbf{w}$). We expect this statistic to fall below the upper 1 - α quantile of the bootstrap distribution.

Technically this means that - unlike Breiman (1992, p.740) and E & T (pp. 113-117) - we do not use the original estimated residuals $\tilde{\mathbf{e}}_r = (\tilde{e}_{r;1}, ..., \tilde{e}_{r;n})'$: some thought shows that when the metamodel is false, then bootstrapping these residuals is wrong! Instead we use the *deviations*.

$$d_{i;r} = w_{i;r} - \bar{w}_i. \tag{8}$$

These variables have zero means, whatever the fitted metamodel is. We use these variables as follows.

First we consider the classic case, namely simulation responses that are independent and have constant variance σ^2. Then we resample *N* values from the *N* original deviations defined in the last equation, (8). This yields the bootstrapped simulation responses assuming the metamodel is valid:

$$w^*_{i;\,r} = \boldsymbol{x_i}\tilde{\beta} + d^*_{i;\,r}\ . \qquad (9)$$

We might refine this bootstrap because only for large replication numbers *m* do the deviations have the same variance as the residuals have, namely σ^2 (i.e., we might multiply $d^*_{i;\,r}$.by an appropriate constant). We point out that no bootstrap simulation response in the last equation (9) is identical to any of the original simulation responses!

The second case allows for variance heterogeneity of the simulation responses. Given the value $\boldsymbol{x}_i$, we then resample only the m_i values of the corresponding $d_{i;\,r}$.

The third and final case allows for common seeds. Then we have *m* independent vectors, each with *n* correlated simulation responses. Now we resample the *m* vectors of deviations $\boldsymbol{d}_r = (d_{r:\,1}, ..., d_{r;\,n})'$ with replacement, so $\hat{F} \sim 1/m$; see (7).

From these bootstrapped I/O data ($\boldsymbol{X},\ \boldsymbol{w}^*$) we compute the EGLS estimators $\boldsymbol{S}^*_w$, $\tilde{\beta}^*$, $\tilde{\boldsymbol{e}}^*$, and F^*; see (3) through (6). Repeating this bootstrap *B* times gives the bootstrap distribution of F^*. If the original statistic does not fall within the interval that ranges from zero to the estimated upper 1 - α quantile of this bootstrap distribution, then we reject the null-hypothesis of a valid metamodel. Note that we keep the original input data ***X*** unchanged; also see Breiman (1992) and E & T.

Note: The pseudorandom number stream may be atypical, resulting in a value *w* that has extremely low probability: *outlier*. For example, the event of a sequence of 1,000 consecutive pseudorandom numbers all below 0.01, is possible, but highly unlikely. Therefore the analysts may wish to eliminate this value *w* when fitting the metamodel. See the general regression literature. In bootstrapping, outliers are automatically removed when they are not resampled.

4 MONTE CARLO STUDY

Consider the best-known queueing model, namely M/M/1, which has Poisson arrival and service processes, one server (implicit assumptions: first-in-first-out or FIFO priority rule, unlimited size of buffer or waiting room). Suppose the simulated response is the steady-state expected waiting time of customers (say) μ. For low traffic rates (say) λ a first-order polynomial is an adequate approximation; for higher traffic rates a second-order polynomial might be adequate (better metamodels are proposed in Cheng and Kleijnen (1998).

However, to improve the *efficiency and effectiveness* of our study on bootstrapping and validation, we proceed as follows. To simulate the M/M/1 model such that accurate estimates of the response parameter μ result, requires extremely long runs. The resulting computer time would be a waste, given the goal of our study. Moreover, we would not know for which traffic rates λ a first or second-order polynomial is adequate. Therefore we make sure that in some Monte Carlo experiments a first-order polynomial with zero intercept ($\beta_0 = 0$) is adequate. In the other experiments we guarantee that the second-order effect is important relative to the first-order effect: see the marginal effect $2\beta_{1;\,1}x_1/\beta_1$. Kleijnen (1992, p. 1172) explains that when estimating the type I errors, the values of the true regression parameters do not matter; so they are all taken equal to zero ($\beta_0 = 0$, $\beta_1 = 0$, ...). We, however, take $\beta_1 = 1$, because of the marginal effect $2\beta_{1;\,1}x_1/\beta_1$. We examine $\beta_{1;\,1}$ is 0.25 in this study.

Here we report on Monte Carlo experiments with a design partly taken from Kleijnen (1992, pp. 1170-3). So we examine a single factor ($k = 1$), a true metamodel that is either a first-order or a second-order polynomial (so *q* is 2 or 3), a number of factor combinations that is small, namely $n = 3$, where the standardized factor values are -1, 0, and 1, a number of simulation replications that is maximal (say) $n + 50 = 53$, true residuals that are either Gaussian or lognormally distributed with zero means, standard deviations $\sigma = (1.0,\ 1.818,\ 0.182)'$, and common seeds which we assume to give $\rho(w_i,\ w_{i'}) = 0.9$, $\alpha = 0.20$, and 100 macroreplications. We use a bootstrap sample size of $B = 1000$, to estimate the distribution of the various validation statistics. (Similar extensive experiments are performed in Breiman 1992).

We first consider normally distributed errors; in this case Rao's lack-of-fit statistic is known to have an $F_{n-q,\ m-n+q}$ distribution in case of a valid metamodel. In the simulation experiments we fit a first-order polynomial to the data, so q= 2. Rao's lack-of-fit statistic is compared to $F_{1,52\ ;\ 0.2}$= 1.6849 and to the 80th percentile of the bootstrap distribution respectively. This allows us tot determine whether the bootstrap gives good results in a analytically tractable case. The simulation results for $\beta_{1;\,1} = 0$ (no specification error) are depicted in Table 1. (A= Accept, R= Reject, B= Bootstrap, F= F-test).

Table 1: $\beta_{1;\,1} = 0$, Normal Errors

B \ F	A	R	T
A	80	1	81
R	1	18	19
T	81	19	100

From these results we estimate and $\hat{\alpha}_F = 0.19$. Under $H_O : \alpha = 0.2$ both results have a p-value of 0.9007 (using and exact binomial test and a two-sided alternative).

To test $H_O : \alpha_B = \alpha_F$ against a two sided alternative only the cases on which they differ are of interest (the R-A and A-R cells in the table). Under H_O each case on which they differ has probability 0.5 of landing in the R-A and A-R cell respectively. We (arbitrarily) define a case in the A-R cell as a "succes", and obtain a binomial test with 2 trials, 1 success and probability of success = 0.5. This gives us a p-value of 1 so we cannot reject the null hypothesis.

Next we consider the case with $\beta_{1;1} = 0.25$ (misspecified model) and normal errors. The results are summarized in Table 2.

Table 2: $\beta_{1;\,1} = 0.25$, Normal Errors

B \ F	A	R	T
A	42	0	42
R	1	57	58
T	43	57	100

The estimated power of the bootstrap in this case is 0.58; and for the F-test 0.57. Clearly we cannot reject H_O : power $_B$ = power $_F$ (p-value = 1).

Next we consider log-normal errors and no specification error $(\beta_{1;1} = 0)$. The simulation results are summarized in Table 3.

Table 3: $\beta_{1;\,1} = 0$, Log-normal Errors

B \ F	A	R	T
A	73	3	76
R	0	24	24
T	73	27	100

From Table 3 we calculate $\hat{\alpha}_B = 0.24$ and $\hat{\alpha}_F = 0.27$, with p-values ($H_O : \alpha = 0.2$) of 0.3176 and 0.1027 respectively. The test of $H_O : \alpha_B = \alpha_F$ yields a p-value of 0.25 on the data in Table 3.

Finally we consider log-normal errors with a misspecified model ($\beta_{1;\,1} = 0.25$).

The simulation results are summarized in Table 4.

Table 4: $\beta_{1;\,1} = 0.25$, Log-normal Errors

B \ F	A	R	T
A	32	5	37
R	0	63	63
T	32	68	100

The estimated power of the bootstrap in this case is 0.63, for the F-test the estimate is 0.68. The test of H_O : power $_B$ = power $_F$ yields a p-value of 0.0625.

5 FUTURE RESEARCH

Popular statistics for measuring the fit of estimated regression metamodels are: the coefficient of determination, denoted as R-square, possibly adjusted for the number of parameters, and the linear correlation coefficient, also known as Pearson's rho. One more alternative is cross-validation, giving Studentized prediction errors and using Bonferroni's inequality.

Several more validation statistics will be considered, such as the relative absolute errors, considering either their average or their maximum. These various fitting and validation procedures and different statistics will be studied through an extensive Monte Carlo experiment. Queueing examples will provide further illustrations of the practical use of these procedures and statistics.

6 CONCLUSIONS

The main conclusions are (i) Not the regression residuals $\hat{e}$ should be bootstrapped; instead the deviations $w - \bar{w}$ should be bootstrapped (ii) bootstrapping Rao's lack-of-fit statistic is a good alternative to the F-test. In case the assumptions of the F-test are known to apply (normal errors), the bootstrap gives virtually identical results. In case the assumptions of the F-test are violated (log-normal errors) both tend to give a somewhat high Type I error probability, but the bootstrap less so. Further experiments are required to draw more definitive conclusions.

ACKNOWLEDGMENT

Cheng and Kleijnen thank the 'NATO Collaborative Research Grants Programme' for the financial support for their joint project on 'Sensitivity analysis for improved simulation modeling'.

REFERENCES

Barton, R.R. and L.W. Schruben (1996), Resampling of empirical distributions for simulation output analysis. Working Paper, PennState, 1996

Breiman, L. (1992), The little bootstrap and other methods for dimensionality selection inregression: x-fixed prediction error. *Journal American Statistical Association*, 87, no. 419, pp. 738-754

Cheng, R.C.H. (1995), Bootstrap methods for computer simulation experiments. *Proceedings of the 1995 Winter Simulation Conference*, edited by C. Alexopoulos, K. Kang, W.R. Lilegdon, and D. Goldsman.

Cheng, R.C.H. and W. Holland (1997), Sensitivity of computer simulation experiments to errors in input data *Journal Statistical Computation and Simulation*, 57, numbers 1-4.

Cheng, R.C.H. and J.P.C. Kleijnen (1998), Improved designs of queueing simulation experiments with highly heteroscedastic responses *Operations Research* (Earlier version: Optimal design of simulation experiments with nearly saturated queues, Discussion Paper, CentER, no. 9567).

Cheng, R.C.H. and J.P.C. Kleijnen (1997) The use of Bayesian methods in regression metamodelling. Working Paper.

Efron, B. and R.J. Tibshirani (1993), *Introduction to the Bootstrap*. Chapman & Hall, New York.

Friedman, L.W. and H.H. Friedman (1995), Analyzing simulation output using the bootstrap method. *Simulation*, 64, no. 2, February 1995, pp. 95-100.

Kim, Y.B., T.R. Willemain, J. Haddock, and G.C. Runger (1993), The threshold bootstrap: a new approach to simulation output analysis. *Proceedings of the 1993 Winter Simulation Conference*, edited by G.W. Evans, M. Mollaghasemi, E.C. Russell, and W.E. Biles, pp. 498-502.

Kleijnen, J.P.C. (1992), Regression metamodels for simulation with common random numbers: comparison of validation tests and confidence intervals. Management Science, 38, no. 8, August 1992, pp. 1164-1185.

Kleijnen, J.P.C. and B. Bettonvil (1990), Measurement scales and resolution IV designs. *American Journal of Mathematical and Management Sciences* 10, nos. 3 & 4, pp. 309-322.

Kleijnen, J.P.C. and Sargent (1997), A methodology for the fitting and validation of metamodels in simulation, Working Paper.

Kleijnen, J.P.C. and Standridge (1987), Experimental design and regression analysis: an FMS case study. *European Journal of Operational Research*, 33, no. 3, pp. 257-261.

Rao, C.R. (1959), Some problems involving linear hypothesis in multivariate analysis. *Biometrika*, 46, pp. 49-58.

Stine, R.A. (1985), Bootstrap prediction intervals for regression. *Journal American Statistical Association*, 80, no. 392, pp. 1026-1031.

AUTHOR BIOGRAPHIES

JACK P.C. KLEIJNEN is Professor of Simulation and Information Systems. His research interests are simulation, mathematical statistics, information systems, and logistics. He published six books and nearly 160 articles; one book was also translated into Russian. He consulted several organizations, and was on many editorial boards and scientific committees. He spent several years in the USA, at universities and companies. He was awarded a number of international fellowships and awards.

AD FEELDERS is an Assistant Professor at the Department of Economics and Business Administration of Tilburg University in the Netherlands. He received his PhD in Artificial Intelligence from the same university, where he currently participates in the Data Mining research program. He worked as a consultant for a Data Mining company. His current research interests include the application of data mining in finance and marketing. His articles appeared in *Computer Science in Economics and Management* and *IEEE Transactions on Systems, Man and Cybernetics*. He is a member of the editorial board of the *International Journal of Intellligent Systems in Accounting, Finance, and Management*.

RUSSELL C. H. CHENG is Professor of Operational Research in the Canterbury Business School at the University of Kent at Canterbury. He has an M.A. from Cambridge University, England. He obtained his Ph.D. from Bath University. He is Chairman of the U.K. Simulation Society, a Fellow of the Royal Statistical Society, Member of the Operational Research Society. His research interests include: variance reduction methods and parametric estimation methods. He is Joint Editor of the *IMA Journal on Mathematics Applied to Business and Industry*, and an Associate Editor for *Management Science.*