Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

REGRESSION METAMODELS AND DESIGN OF EXPERIMENTS

Willem J.H. van Groenendaal
Jack P.C. Kleijnen

Department of Information Systems and Auditing (BIKA)/Center for Economic Research (CentER)
School of Management and Economics (FEW)
Tilburg University (KUB)
5000 LE Tilburg, NETHERLANDS

ABSTRACT

Simulation models are often used to support decision making for problems with uncertain inputs and parameters. Three types of models are used: deterministic, risk, and uncertainty models. Risk models are popular with researchers, but can be used only when the joint probability distribution of the inputs and parameters is known. In many real-life situations, however, this is not the case. Uncertainty models are too restrictive for real-life situations. Therefore deterministic models are then used. The sensitivity of the results is often analyzed by changing one factor at a time or by simulating a few scenarios. This paper, however, shows that in case of uncertainty it might be better to apply design of experiments (DOE) in combination with regression metamodels.

INTRODUCTION

everal methods may be used to analyze the sensitivity of iodel output with respect to changes in input variables, arameters, and model structure. (In the sequel we shall fer to all these changes as factors.) The models we have mind are simulation models that support decision aking when no analytical models are available to escribe the problem and to find an optimal solution. In onomic theory three types of models are distinguished: rtainty or deterministic, risk, and uncertainty models. 'hich type of model is used, depends on what formation is available about the real-life problem. In se of certainty, each set of factor values invariably leads a specific outcome. Economists speak of risk if the ctors are stochastic but their joint probability stribution is known. In case of uncertainty the joint obability distribution is unknown. We restrict certainty here to uncertainty of nature; that is, we clude situations where the managers are competing ainst an intelligent opponent (game theory).

The fact that a model is classified as one of these three types does not mean that the real-life problem falls in the same category. For large econometric and environmental models it is quite common to ignore the stochastic nature of the actual problem, once (part of) the model is estimated. But for relatively simple models, such as net present value (NPV) models, users may also stick to the deterministic analysis. There seems to be a gap between the research community and the other model users, such as businesses and governmental organizations. The former are more interested in risk and uncertainty models, whereas the latter usually stick to deterministic models. We are interested in real-life problems with uncertainty, so we can perform either a risk analysis or a deterministic analysis of the situation.

As case study we use an investment analysis for a gas transmission project on the island of Java in Indonesia (Van Groenendaal, 1996). For the evaluation of this project the World Bank's guidelines are followed (Ward and Deren, 1991). The NPV is the criterion used to judge the investment. This paper aims at methods that can be used within the restricted time available in applied work.

It turns out that in the base-case scenario the project is feasible (NPV = 1,751 billion Rupiahs or 815 million US$ in 1991 prices). Ten factors play a role in the NPV calculation (also see Table 1).

In § 2 we review deterministic, risk, and uncertainty models in relation to robust decision making and their applicability in applied work. In § 3 we use DOE and regression metamodels to obtain the information required by decision makers. In § 4 we present conclusions.

2 ROBUST DECISIONS

When modeling is used to support decision making, the goal is often to support the choice of a policy from a set of alternatives. Particularly in the context of long term decision making (such as large scale investment problems), the uncertainty of future circumstances is a major concern. In practice, often a particular set of values or circumstances is chosen and the problem is analyzed for this set only. The set is either a base-case (most likely), or a worst-case scenario. Based on this analysis, a policy is recommended. What decision makers want to know, is how robust their decision is. "Robust" means that commitments are made as late as possible, and plans can be changed as much as possible when the future is not as anticipated.

Broens (1995, p. 41) uses the term flexible planning as opposed to robust planning. He claims that in robust planning the plan is fixed, but has to cover as many alternative scenarios as possible. This implies that the intent is *not* to change the plan, and to stick to the original commitment. Broens's view is not generally accepted; we use the term robust decision to indicate that a decision has to facilitate as many scenarios as possible, and if possible this decision should include opportunities to alter the plan if circumstances change during execution.

2.1 Deterministic models

When using deterministic models, analysts often rely on some form of what-if analysis to determine the robustness of a decision. Often they change one variable at a time, and they analyze (main) effects. The one-factor-at-a-time approach is popular among economists (Ward and Deren, 1991; Van Groenendaal and Vingerhoets, 1995). This approach, however, is misleading when used to support a decision, because it does not take into account interactions. It is well known among simulationists that it can easily give the decision makers the wrong impression about the problem; in our example no single factor leads to an NPV smaller than zero (§ 3).

In policy related studies, analysts often construct scenarios, that is, a particular choice of factor values. The analysts examine the differences among scenarios. Normally, they compare results with respect to the base case scenario. At best they use a few scenarios; see for example Grübler and McDonald (1996). Managers, however, are often interested in the worst case scenario only. It is difficult to formulate scenarios on which all decision makers agree, especially when the evaluation period is long and the problem complex. The result of the analysis is then weakened by the discussions about the likelihood or even the feasibility of scenarios; alternative scenarios will be proposed. But even if there are no discussions on the scenarios, this approach results only in descriptions of a limited number of futures. In applications often two or three scenarios are used; hardly ever, more than five scenarios are analyzed.

To avoid discussions on scenarios or neglect interactions between factors, we propose to use DOE in combination with regression metamodels. The advantage of this approach is that it gives information on the whole experimental area of interest, to be used by decision makers.

2.2 Risk models

Instead of using a deterministic model to analyze a problem with uncertainty, analysts may introduce a joint probability distribution for the factors. Especially with the growth of software (such as @RISK), this option is becoming more and more popular. Often marginal distributions are formulated based on minimum information per factor: minimum, maximum, and most likely values. The marginal distributions are then combined into a joint distribution, assuming independent factors (for dependent factors see the references in Kleijnen 1996). However, such an approach will not necessarily improve results compared to deterministic models: garbage in garbage out. It does not produce more useful information for the decision makers. As was already shown by Iman and Conover (1982), if information on the joint probability distribution of the factors is wrong, the results obtained from rather arbitrary assumptions on correlations between factors are useless. Furthermore, in many cases the tail of the output distribution is most important for decision support, but this is also the part about which the reliability of the information is lowest; see Kunreuther et al. (1983) for an example.

Instead of using "objective" data, the analysts may use subjective probabilities (expert judgements). Expert judgements, however, will be useful only in case of a limited number of variables and decisions that resemble past decisions. And even then, mistakes are easily made; see Granger and Herion (1990). Again the correlation between factors will be a major problem.

Another approach to getting a robust overall result is to combine the results for a set of scenarios by asking the decision makers to assign subjective probabilities to the individual scenarios (Draper, 1995). These probabilities can then be used to calculate the expected scenario.

Note that in case the information on the joint probability distribution is available but unreliable, the analysts also need to perform a robustness analysis for the distribution, for example, by using the Score Function method; see Rubinstein and Shapiro (1993).

2.3 Uncertainty models

A third type of models is decision models under uncertainty, such as stochastic dominance models (Copeland and Weston, 1992, pp 92-6). This type relies on the axioms of utility theory, which unfortunately are not met in practice (Tversky and Kahneman, 1988).

Most methods using probability or utility are rejected by practitioners. They prefer simple methods, such as one-factor-at-a-time or "switching" values (factor value that causes the NPV to become negative; see Gittinger, 1982, pp 371-3), in combination with the worst case scenario.

Table 1: A Plackett-Burman design for ten factors in the NPV simulation model

combination / factor	1	2	3	4	5	6	7	8	9	10	11	12
1 investment costs	+	+	-	+	+	+	-	-	-	+	-	-
2 construction time	-	+	+	-	+	+	+	-	-	-	+	-
3 reserves West Java	+	-	+	+	-	+	+	+	-	-	-	-
4 real GVA	-	+	-	+	+	-	+	+	+	-	-	-
5 energy prices	-	-	+	-	+	+	-	+	+	+	-	-
6 relative gas/oil price	-	-	-	+	-	+	+	-	+	+	+	-
7 purchase prices	+	-	-	-	+	-	+	+	-	+	+	-
8 coal prices	+	+	-	-	-	+	-	+	+	-	+	-
9 other costs	+	+	+	-	-	-	+	-	+	+	-	-
10 discount rate	-	+	+	+	-	-	-	+	-	+	+	-

For a robust decision the worst case approach is probably the best of the naive methods, because then the actual result will always be better. A problem is how to find the worst case scenario. Moreover, the worst case approach seems rather pessimistic. As Broens (1995) points out, too much attention is given to the worst case; all other information is left unused. Furthermore, the worst case is usually chosen from the likely scenarios; unlikely scenarios are excluded, despite the fact that information on these scenarios might be useful.

A final problem is that risk and uncertainty models often require the problem to fit into a certain form, for example, dynamic programming. Many real-life problems do not meet this requirement.

In summary, many methods for sensitivity analysis do not meet the needs of practitioners. These models are either too rigorous and do not result in the information required by the decision makers, or they require too much information, which is not available. When uncertainty about factors is a real problem and no information on the joint probability distribution can be obtained, DOE combined with regression metamodelling can be a solution, as we shall show next.

3 DESIGN OF EXPERIMENTS

Whatever experimental approach is used, the first problem is to find the appropriate experimental area. Our approach is to start with a base case scenario, that is, a scenario based on the most likely values of the factors. In our investment case study the values concern the amount of gas reserves available in West Java, the growth paths for the Indonesian economy, energy prices, etc.; see Table 1. The base case scenario we obtained through interviews with experts in the various Indonesian ministries and industries involved in the project. For this scenario the project is feasible; that is, it has a positive NPV. Once we had this information, we discussed with our counterparts to what extend the factors could deviate from their base case values. Our analysis focusses on those conditions that will jeopardize the positive advise for the investment project (which follows from the positive value for the NPV in the base case). The reason for this focus is that decision makers are more interested in what can cause the project to become infeasible than in windfall profits. Hence all results in our sensitivity analysis are expected to be worse than the base case result. Identifying the base case values with the 10-dimensional unit vector e_{10} = (+1,+1, ... ,+1), the situation where all factor values deviate from their base case value is (-1,-1, ... ,-1). Opposite to popular believe, (-1,-1, ... ,-1) is not necessarily the worst case scenario, because of interactions; see Tables 1 and 2, combinations 2 and 12 .

Sensitivity analysis should result in information on both main effects (as the one-factor-at-a-time approach does) and interactions. Moreover, we want to execute a minimum number of simulation runs. Therefore we apply a *Plackett-Burman* design (Kleijnen and Van Groenendaal, 1992, pp. 175-7). These designs require a number of runs equal to a multiple of four. Hence, for ten factors a design with twelve runs is used. The design we use is given in Table 1, where + is interpreted as +1 and - as -1; we use standardized factors (see Kleijnen and Van Groenendaal, 1992, pp 177-9). Every column in that table represents the input scenario of a simulation run. It is easily checked that the columns of this design matrix are orthogonal: $(D^T D)^{-1} = 12^{-1} I$ where D is the 12x10 design matrix, the superscript T denotes the transposed matrix, and I denotes the 10x10 identity matrix. Furthermore, the design satisfies one linear constraint: the sum of the first eleven rows of D equals minus row twelve. The augmented matrix $X = (e_{12} : D)$, has the same properties as D has.

When we execute the twelve simulation runs of this design, we get twelve values for the NPV. This is sufficient information to obtain OLS estimates of the main effects only; that is, we fit the first-order metamodel:

$$\underline{y}_i = \sum_{h=0}^{10} \beta_h x_{ih} + \underline{\varepsilon}_i , \tag{1}$$

where $\underline{y}_i$ denotes the response for combination i, x_{ih} are elements of the matrix X, $\underline{\varepsilon}_i$ is an independent and identical distributed error term; and the underscore denotes random variables. However, if interactions are important, then these estimates are biased. In the sequel we assume that there are only two-factor interactions.

Let $\beta_M = (\beta_0, \beta_1, \dots, \beta_{10})^T$ be the vector of coefficients of model (1), and let $\beta_A = (\beta_{1,2}, \beta_{1,3}, \dots, \beta_{9,10})^T$ be the vector of two-factor interactions. Let the matrix of independent variables associated with β_A be denoted by V (with $V = (V_1, \dots, V_9) \in R^{12\times 45}$ and $V_i = (x_i x_{i+1}, \dots x_i x_{10})$ where, e.g., $x_1 x_2$ denotes the 12-dimensional vector with elements $x_{j1}x_{j2}$, j = 1, 2, ... , 12). Then the expected value of $\underline{\beta}_M$ is

$$E\{\underline{\beta}_M\} = \beta_M + 12^{-1} X^T V \beta_A , \tag{2}$$

where $12^{-1}X^TV$ is called the *alias* or bias matrix, which shows how the main effects are confounded with the interactions; see Raktoe, Hedayat, and Federer (1981). Hence the estimator for the main effects would be unbiased if either $X^TV = 0$ or $\beta_A = 0$.

Unfortunately, $X^TV = 0$ does not hold for the Plackett-Burman design. Fortunately this equality can be achieved by applying the *foldover* theorem (see Kleijnen, (1987, p. 303): add $-D$ to the original design matrix D. Hence, twenty-four instead of twelve simulation runs are executed. The resulting design is a resolution IV design; that is, no main effect is confounded with any other main effect or any two-factor interaction; the two-factor interactions, however, are confounded with each other (Kleijnen, 1987, Chapter 25). (Obviously, since there are 1 + 10 + 45 effects and only 24 runs, unbiased estimators of all main effects and two-factor interactions are impossible.) Adding two-factor interactions to (1) and applying the foldover technique leads to

$$\underline{y} = \left(e_{24}, \begin{pmatrix} D \\ -D \end{pmatrix}, \begin{pmatrix} V \\ V \end{pmatrix} \right) \begin{pmatrix} \beta_0 \\ \beta_M \\ \beta_A \end{pmatrix} + \underline{\varepsilon} = Z\beta + \underline{\varepsilon} . \tag{3}$$

This gives the OLS estimator $\beta = (Z^TZ)^{-1} Z^T y$ where

$$Z^TZ = \begin{pmatrix} 24 & 0 & 0 \\ 0 & 24I & 0 \\ 0 & 0 & 2V^TV \end{pmatrix} . \tag{4}$$

Note that the rank of V is eleven, which implies that up to eleven two-factor interactions can be estimated. Next eleven independent columns from V have to be selected, to form the matrix (say) V_I that corresponds with the interactions. The remaining columns of V are combined in the matrix (say) V_A. The resulting alias matrix $(V_I^T V_I)^{-1} V_I^T V_A$ for the eleven interactions can be formed in the same way as we did for $\underline{\beta}_M$ in (2).

Note that if unbiased estimators of more than eleven interactions are needed, the design must be further augmented (beyond the foldover); that is, more simulation runs are required.

Table 2: NPV results (in billions of Indonesian Rupiahs) for the Plackett-Burman design and its foldover

comb.	NPV	comb.	NPV
1	-1,252.8	13	-1,268.0
2	-3,033.8	14	-67.9
3	-1,132.5	15	-1,006.0
4	-1,210.0	16	-1,084.0
5	-1,301.7	17	-1,339.6
6	359.0	18	-2,639.2
7	-997.8	19	-1,362.2
8	454.2	20	-1,353.7
9	-602.6	21	-2,045.6
10	-175.6	22	-1,324.4
11	-1,341.7	23	-278.1
12	-2,985.3	24	1,750.9

The results of the twenty-four simulation runs are shown in Table 2. In this table, combination 13 is the "foldover" of combination 1, ..., combination 24 is the foldover of combination 12 (hence combination 24 is the base case scenario). Because we analyze only factors that cause a decrease in the NPV, there are many negative entries in Table 2. Since each minus value of a single factor lowers the NPV, these negative effects are mitigated only if *interactions* have positive influences. In case the interactions strengthen the main effects, the results of the simulation will be even more negative. The base case gives the NPV result 1,750.9.

The analysis of the data of Table 2 starts with the OLS estimation of the first-order approximation in (1) using the design with 24 runs. The estimates β_0, β_1, β_3, β_5, β_6, β_7, and β_8 are significant at the level $\alpha = 0.05$. The adjusted coefficient of determination R^2_{adj} is 0.88. Because $D^TD = 24I$ (see eq. 4) is a diagonal matrix, the estimates of the main effects do not change when we delete or add main effects.

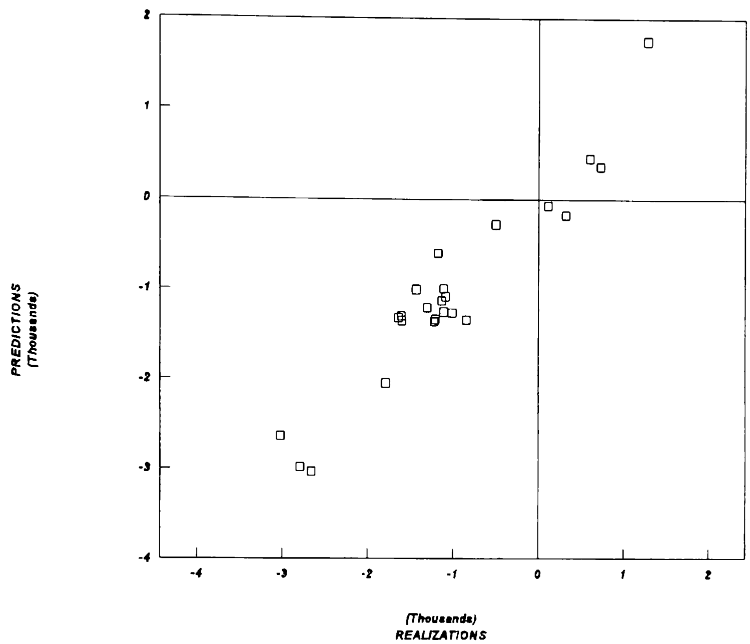

Figure 1: Scatter plot of NPV regression predictions and simulation realizations

Once the significant main effects are known, the significant two-factor interactions may be determined. There are many possible ways to augment (1) with individual interactions. So determining which individual interactions are significant, is problematic, without making further assumptions or executing more simulation runs. A popular assumption is that there are only interactions between factors with *significant* main effects. Estimating this model and testing it against the model without interactions shows that there are indeed some significant interactions: the F-statistic is 3.14, which, however, is barely significant ($F^{6}_{13;0.05} = 2.92$). (Note, that not rejecting H_0 does not prove that there are no interactions.)

Taking into account those interactions that are related to significant main effects is indeed a reasonable approach, when no other information is available (the simulation model is then treated as a black box). However, in our case-study we derive clues from the simulation model itself and from the intermediate simulation results that lead to the outputs in Table 2. We know that in the simulation model, economic growth plays an important role. Nevertheless, Indonesia's economic growth (factor 4) has no significant main effect, β_4. This seems odd, and is also contrary to what economic theory tells us. Studying the detailed simulation results of the twenty-four simulation runs shows that economic growth does strengthen the effect of a change in the West Java reserves (factor 3). Moreover, it strengthens the effects of some of the changes in prices (factors 5, 6, 7, and 8). Therefore we restrict the search for interactions to these six variables (factors 3-8). After testing several alternative specifications of the metamodel, the following model showed the best test results:

$$\begin{aligned}\hat{y} = &-1051.6 + 142.5x_1 + 461.2x_3 + 659.2x_5 + \\ &+ 447.2x_6 + 242.7x_7 + 236.4x_8 + \\ &+ 226.0x_3x_4 + 112.1x_3x_5 - 107.7x_3x_6 + \\ &+ 128.6x_3x_8 .\end{aligned} \tag{5}$$

Its R^2_{adj} is 0.98. The hypothesis $H_0: \beta_{34} = \beta_{35} = \beta_{36} = \beta_{38} = 0$ yields $F^{4}_{13} = 10.81$, which is significant even at the 0.5% level ($F^{4}_{13;0.005} = 6.23$).

The validity of the metamodel (5), relative to the underlying simulation NPV model, can be tested through *cross-validation*: eliminate combinations one by one, re-estimate the regression model, and use that model's prediction $\hat{y}_{-i}$ for the simulation result for the i-th combination eliminated; see Kleijnen and Van Groenendaal (1992, pp. 156-7). To indicate the quality of these predictions $\hat{y}$ we use a scatter plot; see Figure 1. If (5) were perfect, the scatter plot would be a straight line through the origin with an angle of 45°. Actually, the correlation coefficient between $\hat{y}_{-t}$ and y_t is 0.996, which we find satisfactory.

4 CONCLUSIONS

Practitioners often stick to simple methods, such as changing one factor at a time, studying a few scenarios or the worst case scenario only. We think that design of experiments (DOE) in combination with regression metamodelling can be a fruitful method for problems with uncertainties that do not fit into the standard uncertainty models and that have no reliable information on the joint probability distribution. In many areas these circumstances are the rule rather than the exception. Then deterministic models in combination with DOE and regression analysis will provide the decision makers with the information they require. They want to know explicitly what the most important factors are, and how these factors are related. These factors deserve special attention during project execution. Decision makers will use this information, when monitoring the progress of the project, and when designing adjustments whenever progress is not as expected.

We illustrated DOE and regression analysis by a large practical investment problem. It might be argued that our interpretation of interactions is not supported by a complete statistical analysis, since not *all* two-factor interactions were systematically checked. However, analysts who understand their problem will in many cases be able to qualitatively derive which interactions are the most important ones, as we did. The resulting metamodel may be statistically sound and supported by knowledge about the problem at hand. After all, mathematical statistics is only an auxiliary science.

The results obtained for our case-study, using DOE, allow the estimation of the main effects for the ten factors that influence the NPV plus the important interactions between factors. The results have given the Indonesian Government clear insight into the robustness of the project.

REFERENCES

Broens D-F. (1995) *Investment evaluation with respect to commercial uncertainty*, Labyrint Publication, Capelle a/d IJssel, The Netherlands.

Copeland T.E. and J.F. Weston (1992) *Financial theory and corporate policy* (third edition), Addison-Wesley Publishing Company, Reading, Massachusetts, U.S.A.

Draper D. (1995) Assessment and propagation of model uncertainty, *Journal of the Royal Statistical Society*, 57(1):45-97.

Gittinger J.P. (1982) *Economic analysis of agricultural projects*, The John Hopkins University Press, Baltimore and London.

Granger M. M. and M. Henrion (1990) *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*, Cambridge University Press, Cambridge.

Grübler A. and A. McDonald (1996) The drive to cleaner energy, *Options*, IIASA, Laxenburg, Austria.

Iman R.L. and W.J. Conover (1982) A distribution-free approach to inducing rank correlation among input variables, *Communications in Statistics - Simulation and Computation*, 11(3):311-34.

Kleijnen J.P.C. (1987) *Statistical tools for simulation practitioners*, Marcel Dekker, New York, U.S.A.

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

Kleijnen J.P.C. (1996) Five-stage Procedure for the Evaluation of Simulation Models through Statistical Techniques, Proceedings of the 1996 Winter Simulation Conference (edited by J.M. Charnes, D.J. Morrice, D.T. Brunner, and J.J. Swain)..

Kunreuther H.C., J. Linnerrooth, J. Lathrop, H. Atz, S. Macgill, C. Mandl, M. Schwartz, and M. Thompson (1983) *Risk analysis and decision processing: the sitting of liquefied energy gas facilities in four countries*, Springer-Verlag, Berlin, Germany.

Raktoe B.L., A Hedayat, and W.T. Federer (1981) *Factorial designs*, John Wiley & Sons, New York, U.S.A.

Rubinstein R.Y. and A. Shapiro (1993) *Discrete event simulation: sensitivity analysis and stochastic optimization by the score function method*, John Wiley & Sons, Chichester.

Tversky A. and D. Kahneman (1988) Rational choice and the framing of decisions. In DE Bell, H. Raiffa, and A. Tversky (ed.) *Decision making: descriptive, normative, and prescriptive interactions*, Cambridge University Press, Cambridge, Great Britain.

Van Groenendaal W.J.H. and J. Vingerhoets (1995) Can international commodity agreements work? *Journal of Policy Modelling*, 17(3):257-278.

Van Groenendaal W.J.H. (1996) *Investing in energy infrastructures: methodological aspects*, Oxford University Press, Oxford, England. (Forthcoming.)

Ward W.A. and B.J. Deren (1991) *The economics of project analysis: a practitioner's guide*, The World Bank, Washington, D.C.

AUTHOR BIOGRAPHIES

WILLEM J.H. VAN GROENENDAAL is Assistant Professor in the Department of Information Systems and Auditing, and Fellow at the Center for Economic Research (CentER). Both the Department and the Center are within the School of Management and Economics of Tilburg University in Tilburg, The Netherlands. He is

also a consultant for several international development agencies. He has a Ph.D. from Tiburg University. His research interests are in simulation, decision support, and investment analysis.

JACK P.C. KLEIJNEN is Professor of Simulation and Information Systems in the Department of Information Systems and Auditing; he is also associated with the Center for Economic Research (CentER). Both the Department and the Center are within the School of Management and Economics of Tilburg University (Katholieke Universiteit Brabant) in Tilburg, Netherlands. He received his Ph.D. in Management Science at Tilburg University. His research interests are in simulation, mathematical statistics, information systems, and logistics. He published six books and more than 130 articles; lectured at numerous conferences throughout Europe, the USA, Turkey, and Israel; he was a consultant for various organizations; and is a member of several editorial boards. He spent some years in the USA, at different universities and companies. He was awarded a number of fellowships, both nationally and internationally.