faculteit der economische wetenschappen

RESEARCH MEMORANDUM

ILBURG UNIVERSITY
EPARTMENT OF ECONOMICS
ostbus 90153 - 5000 LE Tilburg
etherlands

A Scenario for Sequential Experimentation

Jack P.C. Kleijnen

Professor of Simulation and Information Systems
Department of Information Systems and Accountancy
School of Business and Economics
Catholic University Tilburg (Katholieke Hogeschool Tilburg)
5000 LE Tilburg
Netherlands

September 1985

Key Words and Phrases: Multistage, sequential, screening,
resolution, validation, interactions,
response surface

A SCENARIO FOR SEQUENTIAL EXPERIMENTATION

Jack P.C. Kleijnen

School of Business and Economics
Catholic University Tilburg (Katholieke Hogeschool Tilburg)
5000 LE Tilburg, Netherlands

ABSTRACT

This tutorial emphasizes the role of different types of experimental design in a multi-stage investigation. In the initial phase group-screening can reveal the really important factors among hundreds of factors. Resolution III designs are useful immediately after the screening phase, to investigate first-order effects, provided higher-order effects are unimportant, i.e., validation is necessary. Resolution IV designs may explain why a first-order model is not valid, i.e., they may yield unbiased estimators of sums of interactions. Resolution V designs yield unbiased estimators of the individual two-factor interactions. They can be easily extended to central composite designs to estimate pure quadratic effects of quantitative factors. Smaller steps are also possible, e.g. one run at a time, for model discrimination and calibration.

1. INTRODUCTION

Many papers have appeared discussing new types of designs or refinements of existing designs. In the present paper we assume that the reader is familiar with basic concepts of the design of experiments, such as 2^{k-p} designs. We shall not use technical details of experimental designs. Instead we emphasize the role of different design types in different stages of an investigation. In other words, we assume that we analyze the experimental results of one stage, before we determine the following step in the experiment (feedback principle). Such a multi-stage approach is discussed at length in a number of publications assuming only one or two populations; see the textbooks by Govindarajulu (1981) and Wetherill (1966). We, however, concentrate on $n > 2$ populations, i.e., n combinations of $k > 1$ factors. Each population is replicated $m_i > 1$ times.

To limit the number of references we give one or two publications per item, and refer to Kleijnen (1975, 1986) for more bibliographic data.

We concentrate on experiments with simulated systems. Nevertheless many ideas also apply to experiments with real systems.

2. SCREENING PHASE

In the initial phase of a (simulation) study a large number k of factors are of interest. Then it is too early to use classical designs (like 2^{k-p} designs with $p = 0,1,...$), since classical designs require too many runs $(n > k)$. Moreover, these designs have attractive features like orthogonality, only if we have a firm regression (or Analysis of Variance) model, used to analyze the (simulation) data. For example, in a Plackett - Burman (1946) design the columns of the main effects are orthogonal, i.e., if

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + e_i \qquad (i = 1, \ldots, n) \qquad (1)$$

then this design implies: $\sum_i x_{ij} x_{ij'} = 0$ ($j \neq j'$; $j' = 1, \ldots, k$). However, suppose we discover that a better model is

$$y = \beta_0 + \beta_1 x_1 x_2 + \beta_2 (x_1/x_2) + \ldots + e. \qquad (2)$$

Then the original design results in completely confounded columns for the two new variables ($z_1 = x_1 x_2$ and $z_2 = x_1/x_2$), that is, $x_{1i} x_{2i} = x_{1i}/x_{2i}$ for all i.

Another example is the case study of Kleijnen, Van den Burg and Van der Ham (1979). In this study we simulate part of the Rotterdam harbor, assuming the following effects may be important: six main effects, six specific interactions (β_{12}, β_{23}, β_{24}, β_{25}, β_{26}, β_{13}), one intercept (β_0). First we execute a 2^{6-2} design (with generators 1 = 5 6 and 3 = 4 5) plus ten extra runs for validation. Further study suggests that other interactions, namely β_{14} and β_{15}, might be important. If we had not had ten extra runs, then the 2^{6-2} design (orthogonal in the original thirteen independent variables) would have been very undesirable: the interaction β_{15} would have been completely confounded with the main effect β_6. Fortunately, the ten extra runs result in a non-collinear matrix of independent variables.

For the initial phase of a study we propose <u>group screening</u> designs, based on aggregation; see Kleijnen (1975, 1986). These designs enable us to examine hundreds or thousands of factors in relatively few runs (n << k), provided the number of really important factors (k') is relatively small (k' << k). Other types of design, like random designs, result in biased estimators of the effects (because $n \leq k$). Random designs may be attractive, if we wish to perform a few extra runs to validate the regression model.

It seems dangerous to assume, during the screening phase that no factors interact. Therefore we recommend designs of "resolution IV" or higher during the ($s \geq 2$) stages of the screening phase. In other words, if we examine g group factors then per stage we need a number of runs at least equal to 2g (and rounded upwards to a multiple of eight). At the end of the screening process we have identified the k' important factors (apart from α and β errors).

3. AFTER THE SCREENING

After the screening phase we may feel certain that the (k' important) factors do not interact. At the end of the (s) stages of the screening phase we have observed a total of (say) n runs with k factors; and after those n runs we conclude that only k' (< k) factors are important. Then it may be that these n runs provide a design of high resolution (higher then IV) in these k' factors. Just to illustrate, we consider a case with only four factors examined in a 2^{4-1} design of resolution IV; if the generator was I = 1 2 3 4, and if one factor turns out to be unimportant, then the design is a full factorial for the remaining (three) factors; consequently we can estimate the interactions among these three factors without bias. We can prove that, if in a 2^{k-p} design of resolution r actually no more than r-1 factors are important, then that design is a full factorial in these important factors, possibly with replications; see Kleijnen (1975, pp. 373-374).

Suppose that we have a situation with k factors and k is reasonably small. So we may first have performed a screening experiment (resulting in k important factors; we do not use the symbol k' from here on). Or we may have prior knowledge (for example, an analytic model) that suggests which few factors are important. Should we use a resolution III design or a resolution IV design? We prefer a sequential approach, i.e., we prefer to start with a few observations and to obtain more observations only if needed. We elaborate this approach in the remainder of this paper.

4. RESOLUTION III DESIGNS

We start with a simple regression model, because of the parsimonious character of science. If we reject the original model, then we may use a higher-order model. A higher-order model has more parameters, so that we need more factor combinations. (We might also use transformations or reduce the experimental area; we shall not dwell on these alternatives; see Kleijnen, 1975, 1986.)

The simplest model is a first-order approximation, and it has only k main effects plus the grand mean; see eq. (1). If and only if that model is correct, we obtain unbiased estimators of the k+1 effects using only k+1 - rounded up to the next multiple of four - runs: resolution III design. These designs are tabulated in several publications; see Kleijnen (1975, 1986).

We can test the validity of this simple model, if the number of factor combinations (n) exceeds the number of parameters (k+1); see the cross-validation approach in Kleijnen (1983). Degrees of freedom are indeed available, if k+1 is not exactly a multiple of four. Otherwise we observe some additional combinations (so that n indeed exceeds k+1). Which additional combinations should we observe? We mention:

(i) Combinations which occur in practice.

(ii) The central combination ($x_{ij} = 0$ for $j = 1,\ldots,k$), if the factors are quantitative.

(iii) Combinations used in the debugging of the computer simulation program.

(iv) The "mirror image" of the old design; see next section.

5. RESOLUTION IV DESIGNS

Box and Wilson (1951, p. 35) introduced resolution IV foldover designs, i.e., if we wish a resolution IV design, we double the number of combinations used in the resolution III design, switching signs ($x_{n+i,j} = -x_{i,j}$ with $i = 1,\ldots,n$). Actually, we might execute these additional runs one by one. Then we forecast the response of each new run (using the first-order model calibrated through the resolution III design) and compare the prediction $\hat{y}$ to the actual response y (see Kleijnen, 1983). If we can afford a resolution lV design, then at the end of this stage we have at least k+1 validation runs. There are two situations:

- A model without interactions is valid.
- A model without interactions is not valid.

In the first situation all we have to do is reestimate the effects, using all 2(k+1) runs, where the number of runs may be rounded upwards to the next multiple of eight. In the second situation we expect that some validation runs yield significant forecast errors $(y-\hat{y})$. The resolution IV design may explain why forecast errors occur, because the 2^{k-p} design of resolution IV gives estimates of the sums of two-factor interactions (more precisely this design gives estimates of linear combinations of interactions with weights equal to plus one and minus one; resolution IV designs not equal to 2^{k-p} designs do not yield estimates so simple to interpret; see Kleijnen, 1975, 1986). When the first-order model turns out to be invalid, then we may proceed to a higher-order model. Unfortunately the latter model has many more parameters: besides the grand mean and the k first-order effects there are k(k-1)/2 two-factor interactions, and if (k) factors are quantitative then there are (k) pure quadratic effects. May we eliminate factors before we continue beyond the resolution IV design?

If the factors are qualitative, then the resolution IV design yields unbiased estimators of the main effects (we assume that interactions among three or more factors are zero; such interactions would be hard to interpret). We might assume that factors without main effects have no interactions either (this assumption may be wrong; again see the case study by Kleijnen, Van den Burg, Van der Ham, 1979). Then we eliminate factors with non-significant main effects. It is important to decrease the number of factors (k) when we proceed to a higher-order model, because removing one factor eliminates k-1 two-factor interactions. And dropping a factor may convert the resolution IV design in the original k factors into a resolution V design in the remaining factors (see Section 2).

6. RESOLUTION V AND COMPOSITE DESIGNS

If we use a resolution V design, then we may select either a 2^{k-p} design (see Box and Hunter, 1961) or a Rechtschaffner (1967) design. A 2_V^{k-p} design has one advantage (2_r^{k-p} denotes a 2^{k-p} design of resolution r): it is possible to select a 2_{III}^{k-p} and a 2_{IV}^{k-p} design (in the preceding stages) such that these two designs form part of the (ultimate) 2_V^{k-p} design; see

Daniel (1956, pp. 96-97). However, a 2_{V}^{k-p} design also has a disadvantage: it requires many more observations (n) than there are effects (q): $n \gg 1 + k + k(k-1)/2$. Rechtschaffner's designs are saturated (n = q), but in general they do not comprise the resolution III and resolution IV designs as subsets.

If the factors are quantitative, then the model may contain pure quadratic effects and a resolution V design does not provide unbiased estimators. However, it is easy to augment the resolution V design: we simply add 2k axial points and a few replications of the central point and so we obtain a central composite design; see Draper (1982), Kleijnen (1975, 1986).

7. REFINEMENTS

It is possible to proceed in smaller steps than we outlined above. For example, a 2_{IV}^{k-p} design provides estimates of linear combinations of two-factor interactions. Now suppose that all these combinations of interactions are non-significant except for one combination (say $\beta_{24} + \beta_{18} + \beta_{35} + \beta_{67}$ in a 2^{8-4} design). We might interpret this result as follows: all interactions are zero except for the (four) interactions present in the significant combination; we ignore the pathological case of interactions cancelling out in the non-significant combinations of interactions. Then we do not proceed from the resolution IV design to a resolution V design (namely, the 2^{8-2} design), as the resolution V design yields estimates of all (twenty-eight) interactions whereas we assume that only a few (namely four) interactions are important. There do exist techniques for the selection of small additional fractions. Because it is difficult to formulate simple rules for these special cases, we refer to the literature (Addelman, 1969; Kleijnen, 1975, 1986).

Instead of proceeding in (small or big) steps (multi-stage approach) we may add a single observation at a time (purely sequential approach). Suppose we have already obtained n observations and we maintain M (≥ 2) possible models, for example, a polynomial and an exponential model. If we wish to discriminate among these M alternative models, which observation should we take next? A simple solution is to take the new observation at the point within the experimental area where the two models (assuming M = 2) yield estimated responses furthest apart, i.e., if the two models have parameter vectors β_I and β_{II} then we select x such that $\{y(\hat{\beta}_I, x) - y(\hat{\beta}_{II}, x)\}^2$ is maximal. If there are more than two competing models (M > 2), we may rank the M models using a validation criterion like $\max | y_i - \hat{y}_i |$; next we apply the simple solution to the two best fitting models. For more sophisticated proposals we refer to Hill (1978) and Kleijnen (1986).

We note that in RSM it is also customary to maintain two possible models. However, these two types have a specific relationship: either model I is a first-degree polynomial and model II is a second-degree polynomial, or model I is a second-degree polynomial and model II is a third-degree polynomial. Further most designs are not sequential. The optimal design then minimizes possible bias caused by second-order terms and by third-order terms respectively; see Draper (1982), Kleijnen (1975, 1986).

For quantitative factors we augment the number of levels, as we proceed from a first-order to a second-order model, i.e., in central composite designs the number of levels increases from two to five. For qualitative factors we may also wish to investigate additional factor levels after we have screened the environmental variables, i.e. after this screening we may explore one or more controllable factors in more detail; see Kleijnen (1985).

We may add runs, not to estimate a new model but to obtain more accurate estimators of a given model. We may select these additional runs, maximizing the determinant of X'X; see Johnson and Nachtsheim (1983), Kleijnen (1986).

So there are three situations:
(i) We have rejected model I (for example, a first-order model) and we proceed to a specific model II (for instance, model I augmented with two-factor interactions). How should we augment the experimental design? From a resolution III design we may go to a resolution V design.
(ii) Currently we maintain $M(\geq 2)$ models, for example, a polynomial model and an exponential model. We may select the next observation at the point where the models differ most.
(iii) We trust our final model, and we wish to improve the accuracy of its parameter estimators (calibration). Then we may maximize $|X'X|$.

8. CONCLUSION

If there are very many factors, then we should use group screening. If there are reasonably few factors, then we may start with a resolution III design in the original factors, and estimate the first-order regression model. We validate that model, adding extra runs. If the model is not valid, we can explain this inadequacy, estimating linear combinations of two-factor interactions, through a resolution IV design. We sometimes eliminate factors, and then the resolution IV design may become a design of higher resolution for the remaining factors. We continue with a higher-order model and use a resolution V design to estimate two-factor interactions; if factors are quantitative we augment the design to a central composite design.

REFERENCES

Addelman, S. (1969). Sequences of two-level fractional factorial plans. Technometrics, 11, no. 3: 477-509.

Box, G.E.P. and J.S. Hunter (1961). The 2^{k-p} fractional factorial designs, Part II. Technometrics. 3: 449-458.

Box, G.E.P. and K.B. Wilson (1951). On the experimental attainment of optimum conditions. Journal Royal Statistical Society, Series B, 13, no. 1: 1-38.

Daniel, C. (1956). Fractional replication in industrial research. Proceeding Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5, (J. Neyman, ed.), University of California Press, Berkeley.

Draper, N.R. (1982). Center points in second-order response surface designs. Technometrics, 24, no. 2: 127-133.

Govindarajulu, Z. (1981). The Sequential Statistical Analysis. American Sciences Press, Columbus (Ohio).

Hill, P.D.H. (1978). A review of experimental design procedures for regression model discrimination. Technometrics, 20, no. 1: 15-21.

Johnson, M.E. and C.J. Nachtsheim (1983). Some guidelines for constructing exact D-optimal designs on convex design space. Technometrics, 25, no. 3: 271-277.

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

Kleijnen, J.P.C. (1983). Cross-validation using the t statistic. European Journal Operational Research. 13, no 2: 133-141.

Kleijnen, J.P.C. (1985). On the interpretation of variables. Simulation. 44, no. 5: 237-241.

Kleijnen, J.P.C. (1986). Statistical Tools for Simulation Practitioners. Marcel Dekker, Inc., New York.

Kleijnen, J.P.C., A.J. Van den Burg and R.T. Van der Ham (1979). Generalization of simulation results: practicality of statistical methods. European Journal of Operational Research. 3: 50-64.

Plackett, R.L. and J.P. Burman (1946). The design of optimum multifactorial experiments. Biometrika. 33: 305-325.

Rechtschaffner, R.L. (1967). Saturated fractions of 2^n and 3^n factorial designs. Technometrics. 9: 569-575.

Wetherill, G.B. (1966). Sequential Methods in Statistics. Methuen & Co., Ltd., London, and John Wiley & Sons, Inc., New York. (Revised edition: Chapman and Hill, Ltd., London, 1975.)

IN 1984 REEDS VERSCHENEN

138 G.J. Cuypers, J.P.C. Kleijnen en J.W.M. van Rooyen
Testing the Mean of an Asymetric Population:
Four Procedures Evaluated

139 T. Wansbeek en A. Kapteyn
Estimation in a linear model with serially correlated errors when observations are missing

140 A. Kapteyn, S. van de Geer, H. van de Stadt, T. Wansbeek
Interdependent preferences: an econometric analysis

141 W.J.H. van Groenendaal
Discrete and continuous univariate modelling

142 J.P.C. Kleijnen, P. Cremers, F. van Belle
The power of weighted and ordinary least squares with estimated unequal variances in experimental design

143 J.P.C. Kleijnen
Superefficient estimation of power functions in simulation experiments

144 P.A. Bekker, D.S.G. Pollock
Identification of linear stochastic models with covariance restrictions.

145 Max D. Merbis, Aart J. de Zeeuw
From structural form to state-space form

146 T.M. Doup and A.J.J. Talman
A new variable dimension simplicial algorithm to find equilibria on the product space of unit simplices.

147 G. van der Laan, A.J.J. Talman and L. Van der Heyden
Variable dimension algorithms for unproper labellings.

148 G.J.C.Th. van Schijndel
Dynamic firm behaviour and financial leverage clienteles

149 M. Plattel, J. Peil
The ethico-political and theoretical reconstruction of contemporary economic doctrines

150 F.J.A.M. Hoes, C.W. Vroom
Japanese Business Policy: The Cash Flow Triangle
an exercise in sociological demystification

151 T.M. Doup, G. van der Laan and A.J.J. Talman
The $(2^{n+1}-2)$-ray algorithm: a new simplicial algorithm to compute economic equilibria

IN 1984 REEDS VERSCHENEN (vervolg)

152 A.L. Hempenius, P.G.H. Mulder
Total Mortality Analysis of the Rotterdam Sample of the Kaunas-Rotterdam Intervention Study (KRIS)

153 A. Kapteyn, P. Kooreman
A disaggregated analysis of the allocation of time within the household.

154 T. Wansbeek, A. Kapteyn
Statistically and Computationally Efficient Estimation of the Gravity Model.

155 P.F.P.M. Nederstigt
Over de kosten per ziekenhuisopname en levensduurmodellen

156 B.R. Meijboom
An input-output like corporate model including multiple technologies and make-or-buy decisions

157 P. Kooreman, A. Kapteyn
Estimation of Rationed and Unrationed Household Labor Supply Functions Using Flexible Functional Forms

158 R. Heuts, J. van Lieshout
An implementation of an inventory model with stochastic lead time

159 P.A. Bekker
Comment on: Identification in the Linear Errors in Variables Model

160 P. Meys
Functies en vormen van de burgerlijke staat
Over parlementarisme, corporatisme en autoritair etatisme

161 J.P.C. Kleijnen, H.M.M.T. Denis, R.M.G. Kerckhoffs
Efficient estimation of power functions

162 H.L. Theuns
The emergence of research on third world tourism: 1945 to 1970;
An introductory essay cum bibliography

163 F. Boekema, L. Verhoef
De "Grijze" sector zwart op wit
Werklozenprojecten en ondersteunende instanties in Nederland in kaart gebracht

164 G. van der Laan, A.J.J. Talman, L. Van der Heyden
Shortest paths for simplicial algorithms

165 J.H.F. Schilderinck
Interregional structure of the European Community
Part II: Interregional input-output tables of the European Community 1959, 1965, 1970 and 1975.

IN 1985 REEDS VERSCHENEN

168 T.M. Doup, A.J.J. Talman
A continuous deformation algorithm on the product space of unit simplices

169 P.A. Bekker
A note on the identification of restricted factor loading matrices

170 J.H.M. Donders, A.M. van Nunen
Economische politiek in een twee-sectoren-model

171 L.H.M. Bosch, W.A.M. de Lange
Shift work in health care

172 B.B. van der Genugten
Asymptotic Normality of Least Squares Estimators in Autoregressive Linear Regression Models

173 R.J. de Groof
Geïsoleerde versus gecoördineerde economische politiek in een twee-regiomodel

174 G. van der Laan, A.J.J. Talman
Adjustment processes for finding economic equilibria

175 B.R. Meijboom
Horizontal mixed decomposition

176 F. van der Ploeg, A.J. de Zeeuw
Non-cooperative strategies for dynamic policy games and the problem of time inconsistency: a comment

177 B.R. Meijboom
A two-level planning procedure with respect to make-or-buy decisions, including cost allocations

178 N.J. de Beer
Voorspelprestaties van het Centraal Planbureau in de periode 1953 t/m 1980

178a N.J. de Beer
BIJLAGEN bij Voorspelprestaties van het Centraal Planbureau in de periode 1953 t/m 1980

179 R.J.M. Alessie, A. Kapteyn, W.H.J. de Freytas
De invloed van demografische factoren en inkomen op consumptieve uitgaven

180 P. Kooreman, A. Kapteyn
Estimation of a game theoretic model of household labor supply

181 A.J. de Zeeuw, A.C. Meijdam
On Expectations, Information and Dynamic Game Equilibria