Planning Follow-Up Runs to Screening Experiments

Oddgeir Samset and John Tyssedal
Department of Mathematical Sciences
The Norwegian University of Science and Technology
N-7491 Trondheim, Norway

Abstract

Based on the ideas of Meyer, Steinberg and Box (1996) a method for planning follow-up runs is developed without using a Bayesian framework. Follow-up runs are found by optimising the Kullback-Leibler symmetric information divergence of the predictive densities for the most plausible models resulting from the initial design. As predictive density we use the estimative density for linear models given by Keyes and Levy (1996). This leads to a simple expression for the information criterion which in addition will be 'nearly' additive. It is therefore often enough to only evaluate the information criterion for every single candidate run.

Key words: Kullback-Leibler information; Regression analysis; Sequential experimentation.

1 Introduction

A sequential experimentation is in many situations to be recommended. As we run sequences of small well planned experiments, we learn more and more about the system or process under study. Hence we may build up an ever-increasing foundation for planning new and better experiments to further study the hypothesis we have about the system.

The initial design in a screening situation is typically a non-replicated 2^{k-p} fractional factorial design (Finney, 1945; Box and Hunter, 1961) of resolution III or IV or a Plackett-Burman design (Plackett and Burman, 1946). Such highly fractionated designs do not always lead to an unambiguous conclusion regarding the identification of the active factors. Hence additional runs are often required to resolve these ambiguities.

It is not always trivial how to choose such additional runs that best incorporate the knowledge already gained from the initial design. Box and Hunter (1961) showed how different fractions can be combined to resolve confounded effects in the 2^{k-p} fractional factorial designs. This technique is especially useful for finding unconfounded estimates of all main effects or unconfounded estimates of the main effect and all two-factor interactions of a specific factor. Combining fractions, however, typically results in many additional runs. Daniel (1962) showed how to choose rather few additional runs to resolve strings of two-factor interactions. Such methods are however not easy to generalise to non-geometric Plackett-Burman designs, or other designs with complex alias-patterns. Methods for augmenting non-geometric PB designs to higher resolution based on information about their projection properties are proposed by Lin and Draper (1992) and Hynén (1996).

In general one may want to choose follow-up runs that not only are able to de-alias strings of confounded effects, but also will discriminate between equally plausible models resulting from the initial design. Meyer et al. (1996) presented an appealing method for finding follow-up runs to discriminate between different models that simultaneously give plausible explanations of the observed response values from the initial design. They use a Bayesian methodology (Box and Meyer, 1993) for calculating the posterior probabilities of a set of models given the observed response values. The follow-up runs are then chosen based on maximising the *MD-criterion* (Box and Hill, 1967). This criterion is a weighted sum of the Kullback-Leibler information divergence for discriminating between pairs of the predictive densities for the plausible models. The weights are the product of the posterior probabilities for the respective pair of models given the observed response values.

In this paper we present a way for applying the ideas of Meyer et al. (1996) for choosing follow-up runs using the Kullback-Leibler information without adopting their Bayesian framework. Hence other techniques for identifying the active factors, such as normal and half-normal plots (Daniel, 1959; 1976) or methods based on regression analysis (Hamada and Wu, 1992; Tyssedal and Samset, 1997; Samset, 1998; Samset, 1999), may be used for analysing the initial design. Section 2 gives an outline of the method. In Section 3 we reconsider the *Injection molding example* and the *Reactor example* (Box, Hunter and Hunter, 1978; Meyer et al., 1996). For both these situations the initial designs are 2^{k-p} fractional factorials, but the suggested method can equally well be applied to choose follow-up runs for non-geometric PB designs or other designs with complex alias-patterns.

2 Outline of the Method

A screening of k factors in n runs is performed using an initial design $\boldsymbol{D}$. The n runs result in a vector $\boldsymbol{Y}$ of response values. When searching for the active factors one realises however that there exist m models $M_1, M_2, \ldots, M_m$ that all give a reasonable fit to the data. The design used is not capable of discriminating between these models in the given situation. Thus, additional runs are needed to resolve the ambiguities.

We assume that the response values can be appropriately approximated by a linear function of the input factors. That is, each model M_i is represented by

$$\boldsymbol{Y} = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon} \tag{2.1}$$

with $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma^2)$, and where $\boldsymbol{X}_i$ is the $n \times t_i$ matrix of regressor variables, which here includes an intercept and selected main effect and interaction terms. Using the method of least squares we find estimators

$$\begin{aligned} \hat{\boldsymbol{\beta}}_i &= (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{Y} \\ \hat{\sigma}_i^2 &= (\boldsymbol{Y} - \hat{\boldsymbol{Y}}_i)'(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_i)/(n - t_i) \end{aligned}$$

where $\hat{\boldsymbol{Y}}_i = \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i$. If $\boldsymbol{X}_i$ is not of full rank the Moore-Penrose pseudo-inverse can be used to find a solution to the normal equations.

Let $\boldsymbol{Y}_0$ be the vector of future response values in n_0 new additional runs using a follow-up design $\boldsymbol{D}_0$. Based on model M_i these response values should then be represented by $\boldsymbol{Y}_0 = \boldsymbol{X}_0\boldsymbol{\beta}_i + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma^2)$. The main issue is now to choose these additional runs in order to best discriminate between the m models. Given the vector of response values $\boldsymbol{Y} = \boldsymbol{y}$

from the initial design it thus seems reasonable to choose the follow-up design $\boldsymbol{D}_0$ so that the vector of predicted values $\hat{\boldsymbol{Y}}_{0i} = \boldsymbol{X}_0\hat{\boldsymbol{\beta}}_{0i}$ differ as much as possible for all the m models. One of the models will then hopefully be more consistent with the observed data from the follow-up design compared to the other models.

2.1 Kullback-Leibler Information

Let $f_i(\boldsymbol{y}_0|\boldsymbol{Y}, M_i)$ be the predictive density of $\boldsymbol{Y}_0$ using model M_i given the vector of response values $\boldsymbol{Y} = \boldsymbol{y}$ from the initial design. The *Kullback-Leibler directed divergence* (Kullback, 1968)

$$I(i,j) = \int \ln\left(\frac{f_i(\boldsymbol{y}_0|\boldsymbol{Y}, M_i)}{f_j(\boldsymbol{y}_0|\boldsymbol{Y}, M_j)}\right) f_i(\boldsymbol{y}_0|\boldsymbol{Y}, M_i)\, d\boldsymbol{y}_0 \tag{2.2}$$

measures the information available to discriminate between the predictive densities f_i and f_j. The directed divergence $I(i,j)$ is the expectation $E_{f_i}[ln(f_i/f_j)]$ taken with respect to f_i. Since it is not known whether model M_i or M_j best represents the future response values, both models should be treated equally. A reasonable alternative is thus to use the symmetric divergence given by $J(i,j) = I(i,j) + I(j,i)$.

A good choice of follow-up runs is those who discriminate between all models simultaneously. Thus all values of $J(i,j)$ should be as large as possible for all $i < j$. Meyer et al. (1996) propose to use the MD-criterion given by

$$MD = \sum_{i<j} P(M_i|\boldsymbol{Y})P(M_j|\boldsymbol{Y})J(i,j) \tag{2.3}$$

as a function to optimise. Assume that the posterior probabilities of the models given the response values are all equal. Maximising MD is then similar to maximising the mean symmetric divergence $J_{\text{mean}} = \sum_{i<j} J(i,j)/\binom{m}{2}$. However, J_{mean} may be large even if there only is information to discriminate between two or a rather few of the models. In addition to J_{mean} it therefore seems natural to consider $J_{\text{min}} = \min_{i<j} J(i,j)$, which also should be as large as possible. Both these criteria should be viewed together. A plot of J_{mean} versus J_{min} will give a good illustration of which runs or combinations of runs that perform well according to both criteria.

2.2 Predictive Density

Given an appropriate predictive density one has to choose the design $\boldsymbol{D}_0$ to maximise the Kullback-Leibler symmetric information divergence $J(i,j)$ for all $i < j$. Keyes and Levy (1996) have derived and tabulated nine different predicting densities for the multivariate linear model. We will here use the estimative density, EST(α), given by

$$\boldsymbol{Y}_{0i}|\boldsymbol{Y}, M_i \sim N(\hat{\boldsymbol{y}}_{0i}, \alpha_i\tilde{\sigma}_i^2\boldsymbol{I}) \tag{2.4}$$

where $\tilde{\sigma}_i^2 = \hat{\sigma}_i^2(n - t_i)/n$ is the MLE of σ^2. The scalar α_i is a compensation for the use of the MLE estimators of $\boldsymbol{\beta}_i$ and σ^2, which tend to give a too narrow density. Keyes and Levy propose to use

$$\alpha_i^* = \frac{n}{n_0(n - t_i - 2)}\text{tr}[\boldsymbol{I} + \boldsymbol{X}_{0i}(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_{0i}']. \tag{2.5}$$

See Keyes and Levy (1996) and Murray (1979) for further details on α_i^*.

Let the covariance matrix for EST(α^*) be written as

$$\boldsymbol{V}_{0i} = \alpha_i^* \tilde{\sigma}_i^2 \boldsymbol{I} = \frac{n - t_i}{n_0(n - t_i - 2)} \mathrm{tr}[\boldsymbol{I} + \boldsymbol{X}_{0i}(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_{0i}']\hat{\sigma}_i^2 \boldsymbol{I} = v_i \hat{\sigma}_i^2 \boldsymbol{I}.$$

The general expression for the Kullback-Leibler information divergence between two multivariate normal densities may be found in Kullback (1968)

$$I(i,j) = \frac{1}{2}[\,(\hat{\boldsymbol{y}}_{0i} - \hat{\boldsymbol{y}}_{0j})'\boldsymbol{V}_{0j}^{-1}(\hat{\boldsymbol{y}}_{0i} - \hat{\boldsymbol{y}}_{0j}) + \mathrm{tr}(\boldsymbol{V}_{0j}^{-1}\boldsymbol{V}_{0i}) - \ln(|\boldsymbol{V}_{0i}|/|\boldsymbol{V}_{0j}|) - n_0\,], \tag{2.6}$$

which when the estimative density EST(α^*) is used, simplifies to

$$I(i,j) = \frac{1}{2v_j\hat{\sigma}_j^2}[\,(\hat{\boldsymbol{y}}_{0i} - \hat{\boldsymbol{y}}_{0j})'(\hat{\boldsymbol{y}}_{0i} - \hat{\boldsymbol{y}}_{0j}) + n_0(v_i\hat{\sigma}_i^2 - v_j\hat{\sigma}_j^2) - v_j\hat{\sigma}_j^2 n_0 \ln(\frac{v_i\hat{\sigma}_i^2}{v_j\hat{\sigma}_j^2})\,] \tag{2.7}$$

$$J(i,j) = \frac{1}{2v_iv_j\hat{\sigma}_i^2\hat{\sigma}_j^2}[\,(v_i\hat{\sigma}_i^2 + v_j\hat{\sigma}_j^2)(\hat{\boldsymbol{y}}_{0i} - \hat{\boldsymbol{y}}_{0j})'(\hat{\boldsymbol{y}}_{0i} - \hat{\boldsymbol{y}}_{0j}) + n_0(v_i\hat{\sigma}_i^2 - v_j\hat{\sigma}_j^2)^2\,]. \tag{2.8}$$

Use of EST(α^*) results here in a rather simple expression for the Kullback-Leibler information due to the diagonal covariance matrix. Hence $J(i,j)$ does not require a number of computations of inverses or determinants of matrices, and it is therefore fast to evaluate.

The estimative density EST(α) is probably not the best choice of predictive density. Of the nine different predictive densities considered, Keyes and Levy (1996) recommend to use the Bayesian predictive density (BN) corresponding to the Jeffreys non-informative prior. They point out that this density also can be justified in a non-Bayesian setting since they show it to be optimal using a discrimination information argument. BN is in this situation the density of a multivariate Student-t distribution with expectation $\hat{\boldsymbol{y}}_{0i}$ and covariance matrix $\boldsymbol{W}_{0i} = [\boldsymbol{I} + \boldsymbol{X}_{0i}(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_{0i}']\hat{\sigma}_i^2(n - t_i)/(n - t_i - 2)$. Instead of evaluating the Kullback-Leibler information for the multivariate Student-t density, one could rather use a multivariate normal density with expectation $\hat{\boldsymbol{y}}_{0i}$ and covariance matrix $\boldsymbol{W}_{0i}$. This would however require a number of computations of inverses and determinants of matrices. Observe that the covariance matrices $\boldsymbol{V}_{0i}$ and $\boldsymbol{W}_{0i}$ are equal for $n_0 = 1$. This means that the Kullback-Leibler information is the same for single runs whether one uses a multivariate normal density with covariance matrix $\boldsymbol{V}_{0i}$ or $\boldsymbol{W}_{0i}$.

A problem with both EST(α^*) and BN is the factor $n - t_i - 2$ in the denominator for their respective covariance matrices. Thus, the number of regression variables (including the intercept) t_i has to be less than $n - 2$. This may be a problem if n is as low as 8. We can then at best include only 4 main effect and/or interaction terms in the model.

2.3 Additivity in Kullback-Leibler Information

If Y_{01} and Y_{02} are two independent random variables under both f_i and f_j, there is additivity in the Kullback-Leibler information, i.e. $I(i,j;Y_{01},Y_{02}) = I(i,j;Y_{01}) + I(i,j;Y_{02})$. It is then enough to evaluate the information divergence for each candidate run separately. The information for all possible combinations of different candidate runs can then easily be computed using the additivity property. This will reduce the computing time significantly.

There is additivity in the Kullback-Leibler information using EST(α) for $\alpha_i = 1$. This is in general not true for α_i^* since the marginal predictive densities may be dependent on the number of follow-up runs n_0 due to $\mathrm{tr}[\boldsymbol{I} + \boldsymbol{X}_{0i}(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_{0i}']$. In a designed experiment the 'correlations' between the different contrast columns typically are low, and additivity may therefore often be used as a crude approximation. An impression of the degree of additivity can be obtained by summing the Kullback-Leibler information for all the single runs and compare this to the information computed for all runs simultaneously.

It is here interesting to note that if there is additivity in the information $I(i,j)$, then there is also additivity in $J(i,j)$ and J_{mean}. Hence, it is no need for an exhaustive search for the combination of runs with highest value of J_{mean}. Just choose the n_0 single runs with highest symmetric information divergence. There is of course not additivity in J_{min}. Nevertheless, additivity in $J(i,j)$ implies that any criterion can be used without recomputing the Kullback-Leibler information. An exhaustive search or use of a search algorithm will be useful to evaluate the criterion for all combination of runs.

2.4 Orthogonal Blocking

An other important problem regarding follow-up designs is to decide how many additional runs to include. In this respect it would also be favourable if the follow-up runs constitute an orthogonal block. One should at least include enough follow-up runs to enable the estimation of a possible block-effect.

Since we do not know the correct model we have to consider all the plausible models from the initial design simultaneously. Hence a full orthogonal blocking for all the models may require too many experiments. By looking at the effect of a block factor on the individual estimates of the regression parameters for each model, one may nevertheless get useful insight regarding the number of runs to include. One possibility is to study the variance inflation factor (Marquardt, 1970) of all terms for each model both with and without the block factor included. The variance inflation factor (VIF) gives the increase in variance for the individual estimated regression parameters due to linear dependencies between the regression variables in a model compared to if orthogonality exists, i.e. $\text{VIF}(\hat{\beta}_j) = 1/(1 - R_j^2)$ where R_j^2 is the multiple coefficient of determination when regressing the variable x_j on the other regression variables in the model. It is therefore possible to see which and to what extent the estimated regression parameters will be influenced by the inclusion of a block factor. When there are many combinations of runs giving the same information criterion, use of VIF can be helpful for choosing the best of these combinations with respect to orthogonal blocking. If the number of candidate runs is large it may however be time consuming to compute and examine all possibilities.

Since the follow-up runs are based on the somewhat limited knowledge from the initial runs, there are also some limitations in the knowledge to extract about the candidate runs and their ability to discriminate between the different models. It therefore seems reasonable to have a sequential assembly of follow-up runs if this is possible. That is, it may be better to run more small blocks than one large block of runs, updating the models and the knowledge after each new follow-up design.

3 Examples

3.1 Injection Molding Example

We here reconsider the *Injection molding example* by Box et al. (1978), which was used by Meyer et al. (1996) to demonstrate the usefulness of the MD-criterion for planning follow-up designs. A 2^{8-4}_{IV} fractional factorial design with generators $I = ABDH = ACEH = BCFH = ABCG$ was initially employed for screening eight factors in 16 runs. From the normal probability plot the estimated contrasts corresponding to the main effects C and E together with the string of aliased two-factor interactions $AE + BF + CH + DG$ stand out as distinguishable from noise. The design, response values and normal probability plot of the estimated contrasts can be found in both of the references given above.

By using the method of Box and Meyer (1993), Meyer et al. (1996) found that the five models characterised by the factors (A, C, E), (A, C, H), (A, E, H), (C, E, H) and (A, C, E, H) account for 99.9% of the total posterior probability. These models include all corresponding main effect, two- and three-factor interaction terms. Finally, the 10 best follow-up designs with four runs according to the MD-criterion were given.

In Table 1 we have used the five models found by Meyer et al. but where the three-factor interactions are neglected. These five models together hold four potentially active factors, and follow-up runs are thus chosen among the runs from the full 2^4 design. Eight of the runs from this 2^4 design were used in the initial design for the four potentially active factors. First J_{mean} and J_{min} are listed when including only the 8 runs used in the initial design, only the 8 runs not used, and all 16 runs together. Observe that J_{mean} for the runs used in the initial design added to J_{mean} for the runs not used is approximately the same as J_{mean} for all runs together. This indicates that the symmetric information is approximately additive here when $\text{EST}(\alpha^*)$ is used as the predictive density. For the 8 runs not used in the initial design J_{mean} is about twice as large as for the runs already used. This is however very different for J_{min} which is 0.92 and 17.57 for used and not used runs, respectively.

We have further listed the information criteria (ranked according to J_{mean}) for each of the 8 candidate runs not used in the initial design. The runs are numbered according to the 2^4 design given in standard order. The numbering of the runs used by Meyer et al. is given in parentheses. The sum of J_{mean} for these 8 runs considered individually is equal to J_{mean} for the 8 runs simultaneously. Assuming additivity in the symmetric information divergence thus seems reasonable here. Finally the six best follow-up designs with two, three and four runs ranked by J_{mean} are listed. These designs are found by examining all possible combinations of the 8 runs not included in the initial design. Repeated runs are not allowed since the follow-up designs with highest J_{mean} then would contain only repeated runs due to the additivity in the information divergence. The design with four runs having highest J_{mean} contains run 3, 9, 12 and 15. These runs correspond to the four runs 11, 9, 15 and 12 given by Meyer et al., which is the follow-up design ranked as number three according to the MD-criterion. Note that Meyer et al. select runs among all 16 runs also allowing for repeated runs. The five best follow-up designs with four runs according to the MD-criterion without repeated runs are among the six best designs given here. These results show, at least for this situation, that there is a fairly good correspondence between the follow-up designs chosen by the MD-criterion and the method presented here when the models to discriminate are almost the same.

In Figure 1 the values of J_{min} are plotted versus the values of J_{mean} for all follow-up

Table 1: *Injection Molding Example*: Kullback-Leibler symmetric information divergence for discriminating the five models from Meyer et al. (1996).

```
Models to discriminate:
=======================
M1 :  s= 1.07  1 3 5 1*3 1*5 3*5
M2 :  s= 2.63  1 3 8 1*3 1*8 3*8
M3 :  s= 3.73  1 5 8 1*5 1*8 5*8
M4 :  s= 0.85  3 5 8 3*5 3*8 5*8
M5 :  s= 0.75  1 3 5 8 1*3 1*5 1*8 3*5 3*8 5*8

Potentially active factors: 4
  (1 3 5 8)

Kullback-Leibler Information:
=============================
            No.   J_mean    J_min
Used          8    40.10     0.92
Not used      8    78.98    17.57
All          16   118.29    24.84

Rank    J_mean    J_min    Run
   1     12.99     3.96    15 (12)
   2     12.91     2.87     9 ( 9)
   3     11.55     2.62    12 (15)
   4      9.90     1.52     3 (11)
   5      8.52     1.32    14 (14)
   6      8.36     0.31     8 (16)
   7      7.68     0.31     2 (13)
   8      7.07     0.98     5 (10)
Sum:     78.98
```

```
Combinations of runs (exhaustive search):
=========================================
 Rank    J_mean    J_min    Run
    1     25.90     8.67     9  15
    2     24.54     6.95    12  15
    3     24.45     6.63     9  12
    4     22.89     5.86     3  15
    5     22.81     5.86     3   9
    6     21.52     5.29    14  15

    1     37.44    11.29     9  12  15
    2     35.80    10.19     3   9  15
    3     34.44     8.47     3  12  15
    4     34.42    11.29     9  14  15
    5     34.36     8.47     3   9  12
    6     34.25     8.98     8   9  15

    1     47.35    12.81     3   9  12  15
    2     45.97    13.90     9  12  14  15
    3     45.80    11.60     8   9  12  15
    4     45.13    11.60     2   9  12  15
    5     44.51    12.81     5   9  12  15
    6     44.33    12.81     3   9  14  15
```

designs with four runs selected among all 16 runs (x) or only among the 8 runs not in the initial design (o). Repeated runs are not considered. We observe that designs with high values of J_{mean} also tend to have relatively high values for J_{min}. Thus, in this situation J_{mean} gives good follow-up designs also with respect to J_{min}.

The follow-up design found in Box et al. (1978) using the method of Daniel (1962), is ranked as number 1242 according to the MD-criterion. The large estimated contrasts, corresponding to the two main effects and the string of four aliased two-factor interactions, imply that there are potentially 15 different models to discriminate between. A somewhat simpler problem is to discriminate between the four models in Table 2. Note that if this can be done, the string of confounded two-factor interactions will be disentangled. These four models together contain eight potentially active factors. Hence the follow-up runs are now chosen among the runs from the full 2^8 design not included in the initial design. Observe that the 10 follow-up designs with one run listed in Table 2 all have the same value $J_{\mathrm{mean}} = 6.69$ and $J_{\mathrm{min}} = 0.00$. Table 3 shows however that there are three main 'types' of runs. The 96 runs ranked as number 1 to 96 are all equivalent in the sense that they have the same values of J_{mean} and J_{min}. All these runs will individually enable to discriminate between two and two models. The 128 runs ranked as number 97 to 224 only discriminate one model from the three other. The last 14 runs ranked as number 225 to 240 do not discriminate among any of the four models. We see that run 4 discriminates between the two groups of models (M_1,M_2) and (M_3,M_4), but can not discriminate between the models within each group. Similarly run 6 discriminates between the two groups of models (M_1,M_3) and (M_2,M_4). Thus, run number 4 and 6 are together sufficient for discriminating between these four models and thus

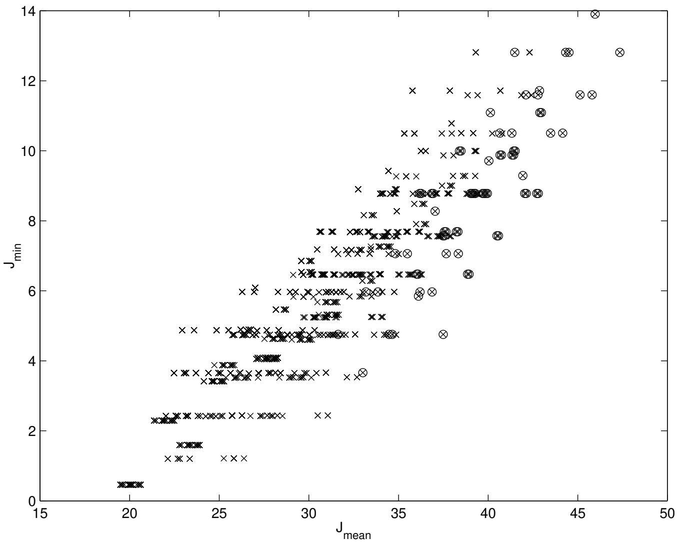

Figure 1: *Injection Molding Example*: $J_{\min}$ versus J_{mean} for follow-up designs with four runs when selection is among all runs (x) and runs not used in the initial design (o).

Table 2: *Injection Molding Example*: Kullback-Leibler symmetric information divergence for discriminating between the four models, i.e. disentangling the confounded two-factor interactions from Box et al. (1978).

```
Models to discriminate:
=======================
M1 :  s= 1.19  3 5 1*5
M2 :  s= 1.19  3 5 2*6
M3 :  s= 1.19  3 5 3*8
M4 :  s= 1.19  3 5 4*7

Potentially active factors: 8
  (1 2 3 4 5 6 7 8)

Kullback-Leibler Information:
=============================
            No.   J_mean    J_min
Used         16     0.00     0.00
Not used    240  1283.64  1283.64
All         256  1283.64  1283.64
```

```
Rank     J_mean    J_min    Run
   1       6.69     0.00      4
   2       6.69     0.00      6
   3       6.69     0.00      7
   4       6.69     0.00     10
   5       6.69     0.00     11
   6       6.69     0.00     13
   7       6.69     0.00     19
   8       6.69     0.00     21
   9       6.69     0.00     24
  10       6.69     0.00     25
 ...
 Sum:   1283.64
```

Table 3: *Injection Molding Example*: Discriminating between the four models, i.e. disentangling the confounded two-factor interactions from Box et al. (1978).

Rank	Run	1 2 3 4 5 6 7 8	y1	y2	y3	y4	J(1,2)	J(1,3)	J(1,4)	J(2,3)	J(2,4)	J(3,4)
1	4	+ + - - - - - -	16.6	16.6	21.2	21.2	0.00	10.03	10.03	10.03	10.03	0.00
2	6	+ - + - - - - -	22.1	26.7	22.1	26.7	10.03	0.00	10.03	10.03	0.00	10.03
3	7	- + + - - - - -	26.7	22.1	22.1	26.7	10.03	10.03	0.00	0.00	10.03	10.03
...												
96	253	- - + + + + + +	18.3	18.3	22.9	22.9	0.00	10.03	10.03	10.03	10.03	0.00
97	2	+ - - - - - - -	16.6	21.2	21.2	21.2	10.03	10.03	10.03	0.00	0.00	0.00
...												
224	255	- + + + + + + +	18.3	22.9	22.9	22.9	10.03	10.03	10.03	0.00	0.00	0.00
225	16	+ + + + - - - -	22.1	22.1	22.1	22.1	0.00	0.00	0.00	0.00	0.00	0.00
...												
240	241	- - - - + + + +	12.8	12.8	12.8	12.8	0.00	0.00	0.00	0.00	0.00	0.00
54	143	- + + + - - - +	26.7	22.1	26.7	22.1	10.03	0.00	10.03	10.03	0.00	10.03
56	148	+ + - - + - - +	17.4	12.8	12.8	17.4	10.03	10.03	0.00	0.00	10.03	10.03
85	227	- + - - - + + +	21.2	21.2	16.6	16.6	0.00	10.03	10.03	10.03	10.03	0.00

enable separate estimation of the four fully confounded two-factor interactions if we neglect any block-effect.

Table 2 and Table 3 show the danger of only focusing on either J_{mean} or J_{min} as a measure to optimise. Any combination of three runs from the 96 runs ranked first in Table 3 has the same value of J_{mean} but can have the three values 0, 10.03 or 20.06 for J_{min}. It is therefore useful to consider both J_{mean} and J_{min}, and if possible also the individual values of $J(i,j)$, together.

Table 3 also lists three of the runs (143, 148 and 227) for the follow-up design actually performed by Box et al. This shows that the proposed follow-up design, neglecting the extra run added to estimate the block-effect, actually is a very good choice for disentangle the string of two-factor interactions in the given situation. It is also interesting to see that these runs correspond to the other follow-up designs with three runs found using the Kullback-Leibler symmetric information divergence in this situation.

The observed response values for the three additional runs are 29.4, 13.6 and 19.7, respectively. By comparing these observed response values to the predicted values given in Table 3 for the four models we see that run 143 strongly supports models M_1 or M_3, run 148 supports model M_2 or M_3 while the observed value of run 227 tends to be located between the predicted values and thus only weakly supports M_1 or M_2.

The four models enable here separate estimation of the four fully confounded two-factor interactions. This does not imply that one of these four models necessarily is the best. One possibility, which should be examined after performing the follow-up design, is whether two or more of the confounded two-factor interactions should be in the model. Specifying a more detailed models when planning the follow-up designs, including more than one two-factor interaction, will however make only minor difference here due to the fully confounded contrast columns.

For any number of runs there are here a number of follow-up designs judged equally well according to J_{mean} and J_{min}. In Table 4 we have given four possible follow-up designs with four runs. The three designs (4,6,28,30), (4,13,21,28) and (4,6,19,21) will all maximise both J_{mean} and J_{min}. The design (143,148,227,256) is the design actually performed by Box et al. The latter design has a slightly smaller value of J_{mean}. For these four designs we have also

Table 4: *Injection Molding Example*: Four possible follow-up designs with four runs.

Runs	VIF	Runs	VIF
4 6 28 30 $J_{\text{mean}} = 26.74$ $J_{\text{min}} = 20.06$	1.00 1.00 1.04 1.04 1.56 1.00 1.04 1.04 1.56	143 148 227 256 $J_{\text{mean}} = 20.06$ $J_{\text{min}} = 20.06$	1.04 1.00 1.00 1.04 1.87 1.00 1.00 1.25 1.87
	1.00 1.04 1.00 1.04 1.56 1.04 1.00 1.04 1.56		1.00 1.00 1.00 1.00 1.56 1.00 1.00 1.00 1.56
	1.00 1.04 1.00 1.04 1.56 1.04 1.00 1.04 1.56		1.00 1.04 1.00 1.04 1.56 1.04 1.00 1.04 1.56
	1.00 1.00 1.04 1.04 1.56 1.00 1.04 1.04 1.56		1.00 1.00 1.04 1.04 1.56 1.00 1.04 1.04 1.56
4 13 21 28 $J_{\text{mean}} = 26.74$ $J_{\text{min}} = 20.06$	1.00 1.00 1.00 1.00 1.56 1.00 1.00 1.00 1.56	4 6 19 21 $J_{\text{mean}} = 26.74$ $J_{\text{min}} = 20.06$	1.04 1.00 1.00 1.04 1.87 1.00 1.00 1.25 1.87
	1.00 1.04 1.00 1.04 1.56 1.04 1.00 1.04 1.56		1.00 1.04 1.00 1.04 1.56 1.04 1.00 1.04 1.56
	1.00 1.04 1.00 1.04 1.56 1.04 1.00 1.04 1.56		1.00 1.04 1.00 1.04 1.56 1.04 1.00 1.04 1.56
	1.00 1.00 1.00 1.00 1.56 1.00 1.00 1.00 1.56		1.04 1.00 1.00 1.04 1.87 1.00 1.00 1.25 1.87

given the VIF for the regression variables for each of the four models both with (first line) and without (second line) a block factor included. Each line give the VIF for the intercept term, the two main effect terms, the interaction term and the block term, respectively, in the given order. We see that for all designs the intercept is affected by the block factor. For design (4,6,28,30) and (4,13,21,28) none of the main effects nor the two-factor interaction are affected by the block. For design (143,148,227,256) the two-factor interaction for model 1 is affected by the block. Actually the inclusion of the block factor increases the variance inflation factor from 1.04 to 1.25, i.e. an increase of 20% in the variance for the corresponding estimated regression parameter β_{15}. This is not too bad, but is possible to avoid in this situation by rather select design (4,6,28,30). For design (4,6,19,21) two of the two-factor interactions will be affected by the block.

3.2 Reactor Example

Meyer et al. (1996) also used the *Reactor example* from Box et al. (1978) to demonstrate the performance of the MD-criterion for selecting follow-up runs. This example is pleasing because all response values for the full 2^5 factorial design are available. By using only the 1/4 fraction of this design, a 2^{5-2}_{III} fractional factorial design with generators $I = ABD = ACE$, Meyer et al. demonstrate that four additional runs based on their MD-criterion enable identifying the active factors. The design and the response values are given in Box et al. (1978).

Figure 2 shows the normal and half-normal plot for both the full 2^5 factorial (top) and the 2^{5-2}_{III} fractional factorial design (bottom). The plots from the full factorial design show that there are five active effects, the main effects of the factors 2, 4 and 5 and the two-factor interaction between factor 2 and 4 and between factor 4 and 5. The plots from the 1/4 fraction are much more difficult to interpret, but indicate four or possibly five active contrasts. Due to the alias-pattern of the design, however, there are a number of models that may give a plausible explanation of the data. Assuming three-factor and higher order interactions to be negligible there are eight plausible models with four terms (in addition to the intercept) that

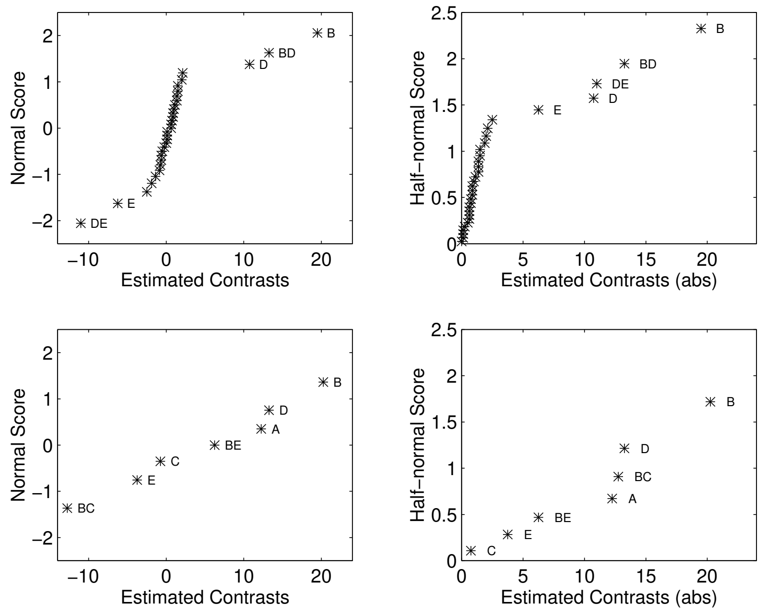

Figure 2: *Reactor Example*: Normal and half-normal plot of the estimated contrasts for the full 2^5 factorial design (top) and the 2^{5-2}_{III} fractional factorial design with generator $I = ABD = ACE$ (bottom).

Table 5: *Reactor Example*: Alias pattern for the 2^{5-2}_{III} fractional factorial design with the eight most plausible models with four terms for explaining the observed response values.

Alias pattern	Most plausible models
$A \to A + BD + CE$	M_1: 1 2 1*2 2*3
$B \to B + AD$	M_2: 1 4 1*4 4*5
$C \to C + AE$	M_3: 2 4 2*3 2*4
$D \to D + AB$	M_4: 2 4 2*4 4*5
$E \to E + AC$	M_5: 1 2 4 2*3
$BC \to BC + DE$	M_6: 1 2 4 4*5
$BE \to BE + CD$	M_7: 2 1*2 2*3 2*4
	M_8: 4 1*4 2*4 4*5

follow the *principle of heredity*, i.e. models where an interaction is included only if at least one of the corresponding factor main effects also is in the model. These models are given in Table 5 together with the alias- pattern of the design. Model M_1 to M_4 are especially convincing since there are only three active factors. Model M_5 and M_6 contain four active factors but are nevertheless acceptable due to a low number of two-factor interactions. Model M_7 and M_8 are maybe less plausible due to relatively many two-factor interactions.

Table 6 shows the Kullback-Leibler symmetric information divergence for discriminating between these eight models containing one to five follow-up runs ranked by J_{mean} (left) and by J_{min} (right). Note that only one design for different values of J_{mean} and J_{min} is listed. The best designs with one run (4) and two runs (4, 10) are the same ranked by either J_{mean}

Table 6: *Reactor Example*: Kullback-Leibler symmetric information divergence for discriminating between the eight models. Follow-up designs with one to four runs using exhaustive search ranked by J_{mean} (left) and J_{min} (right).

Models to discriminate:	Kullback-Leibler Information:			
M1 - M8		No.	J_mean	J_min
	Used	8	0.00	0.00
Potentially active factors: 5	Not used	24	39.40	13.76
(1 2 3 4 5)	All	32	39.40	13.76

Exhaustive search - ranked by mean:				Exhaustive search - ranked by min:			
Rank	J_mean	J_min	Runs	Rank	J_mean	J_min	Runs
- 1	4.30	0.00	4	- 1	4.30	0.00	4
3	2.37	0.00	8	3	2.37	0.00	8
7	2.32	0.00	17	7	2.32	0.00	17
9	1.73	0.00	5	9	1.73	0.00	5
11	1.50	0.00	24	11	1.50	0.00	24
- 1	8.60	1.72	4 10	- 1	8.60	1.72	4 10
2	6.67	0.86	4 14	2	6.67	0.86	4 14
6	6.67	0.01	4 8	6	3.87	0.86	8 30
10	6.62	0.01	4 27	10	3.81	0.86	1 27
12	6.62	0.00	4 17	14	6.03	0.73	4 5
- 1	10.97	1.86	4 8 10	- 1	10.92	2.58	4 10 17
5	10.92	2.58	4 10 17	3	10.33	2.58	4 5 10
7	10.33	2.58	4 5 10	5	10.10	2.58	4 10 24
9	10.10	2.58	4 10 24	7	8.93	2.58	4 17 27
11	10.09	1.86	1 4 10	9	8.35	2.58	4 5 27
- 1	13.34	1.86	4 8 10 14	- 1	13.23	3.44	4 10 17 27
7	13.29	2.80	4 8 10 17	2	12.65	3.44	4 5 10 17
15	13.23	3.44	4 10 17 27	6	12.42	3.44	4 10 17 24
16	12.70	2.80	4 5 8 10	10	12.06	3.44	4 5 10 15
24	12.65	3.44	4 5 10 17	11	11.83	3.44	4 5 10 24

and J_{min}. Observe however that the design with three runs ranked as number five by J_{mean} (4, 10, 17) has a slightly less value of J_{mean} but a relatively much larger value of J_{min} than the four designs ranked above. This is even more clear for the designs with four runs. Hence both J_{mean} and J_{min} should be view together. See also Figure 3 where the values for J_{min} are plotted versus J_{mean} for follow-up designs with four runs.

In Table 7 we have listed the VIF for the eight models for the two follow-up designs (4, 10, 17, 27) and (4, 5, 10, 17). Both these designs have the same values for J_{mean} and J_{min}. However, for the first design all four effects can be estimated unconfounded with the block-effect. This is not true for the latter design.

Since the observed response values are available for the full 2^5 design one has the possibility to examine how good the proposed follow-up designs perform for discriminating between the eight models. In Figure 4 the values of s and R^2 are given when regressing the eight model to the data from the initial design and adding one to four additional runs (4, 10, 17 and 27). The values for s and R^2 are also given when including all 32 runs. A possible block-effect is not accounted for in these calculations. The ability to discriminate between these models is seen to be very good already for 4 additional runs compared to 24 additional runs.

From the normal and half-normal plot of the estimated contrasts in Figure 2 one might argue, or at least take this possibility into consideration, that there are five rather than four active contrasts. Thus, by using the alias-pattern in Table 5 one can specify twelve plausible models that follows the principle of effect heredity. The proposed follow-up designs

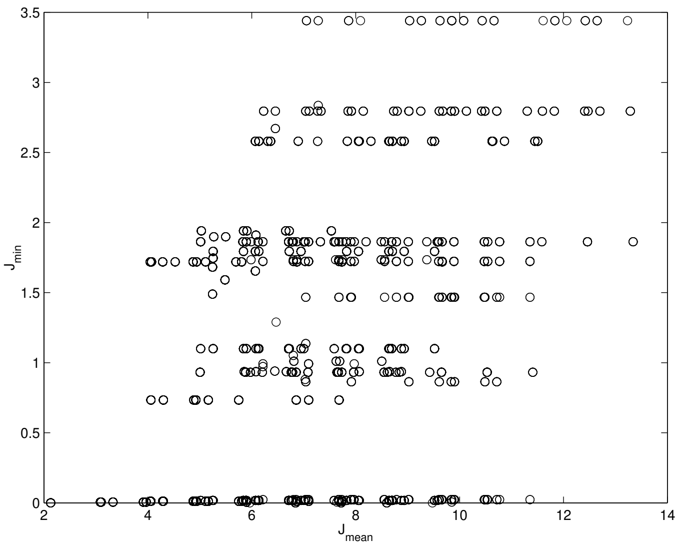

Figure 3: *Reactor Example*: $J_{\min}$ versus J_{mean} for follow-up designs with four runs when selection is among runs not used in the initial design.

Table 7: *Reactor Example*: Two possible follow-up designs with four runs.

Runs	VIF	Runs	VIF
4 10 17 27 $J_{\text{mean}} = 13.23$ $J_{\min} = 3.44$	1.00 1.00 1.12 1.00 1.12 1.12 1.00 1.12 1.00 1.12 1.12	4 5 10 17 $J_{\text{mean}} = 13.23$ $J_{\min} = 3.44$	1.06 1.06 1.09 1.09 1.06 1.29 1.07 1.17 1.17 1.06 1.29
	1.00 1.00 1.00 1.12 1.12 1.12 1.00 1.00 1.12 1.12 1.12		1.06 1.06 1.09 1.09 1.06 1.29 1.07 1.17 1.17 1.06 1.29
	1.00 1.12 1.00 1.12 1.00 1.12 1.12 1.00 1.12 1.00 1.12		1.06 1.09 1.09 1.06 1.06 1.29 1.17 1.17 1.06 1.07 1.29
	1.00 1.12 1.00 1.00 1.12 1.12 1.12 1.00 1.00 1.12 1.12		1.06 1.09 1.09 1.06 1.06 1.29 1.17 1.17 1.07 1.06 1.29
	1.00 1.00 1.12 1.00 1.12 1.12 1.00 1.12 1.00 1.12 1.12		1.06 1.06 1.09 1.09 1.06 1.29 1.07 1.17 1.17 1.06 1.29
	1.00 1.00 1.12 1.00 1.12 1.12 1.00 1.12 1.00 1.12 1.12		1.06 1.06 1.09 1.09 1.06 1.29 1.07 1.17 1.17 1.06 1.29
	1.00 1.12 1.00 1.12 1.00 1.12 1.12 1.00 1.12 1.00 1.12		1.06 1.09 1.09 1.06 1.06 1.29 1.17 1.17 1.06 1.07 1.29
	1.00 1.00 1.12 1.00 1.12 1.12 1.00 1.12 1.00 1.12 1.12		1.06 1.09 1.09 1.06 1.06 1.29 1.17 1.17 1.07 1.06 1.29

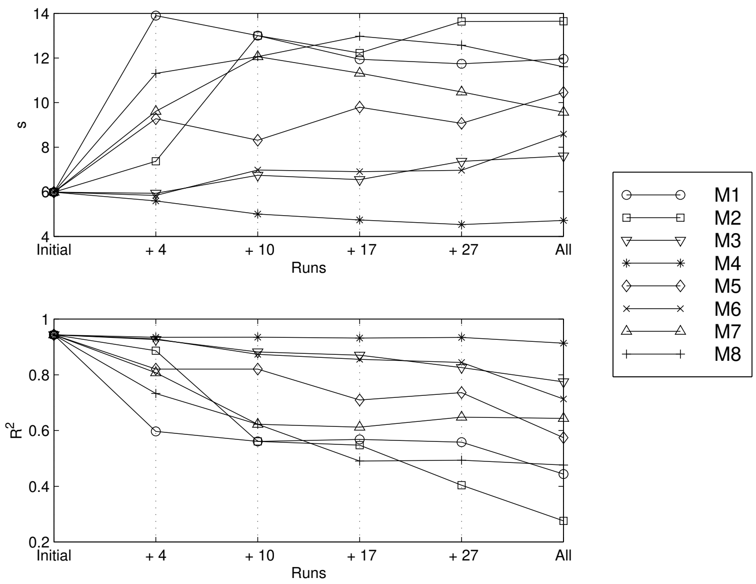

Figure 4: *Reactor Example*: Estimated standard deviation s and R^2 for the 8 models using the observed response values for the proposed follow-up designs.

for discriminating between the twelve rather than the eight models are the same for both one, two, three and four runs ranked by J_{min}. The only difference is now that the follow-up designs ranked by J_{mean} correspond more to the designs ranked by J_{min}. Note that for computing the Kullback-Leibler information for discriminating between the twelve models with five terms, we used $\alpha = 1.0$ since α^* would give division by zero when computing the covariance matrices for the predictive density due to the relatively large number of terms in the models.

4 Concluding Remarks

Based on the ideas from Meyer et al. (1996) a method for planning follow-up designs is developed without a Bayesian framework. This implies that there is no need for finding the posterior probabilities of models given the observed response values from the initial design. Instead of the MD-criterion we rather use the Kullback-Leibler symmetric information divergence. This is however similar to the MD-criterion when all models have the same posterior probability. Without the Bayesian framework other methods for identifying plausible models from the initial design, such as normal and half-normal plots and methods based on regression analysis, can be used. Hence there is also a greater flexibility in how to specify the most plausible models.

To compute the Kullback-Leibler information we have used the estimative density for linear models given by Keyes and Levy (1996). The simple form of this predictive density leads to a rather simple formula for the symmetric information divergence not requiring any computation of inverses or determinants for matrices. In addition this predictive density also leads to near additivity in information divergence which implies that it often is enough to compute the Kullback-Leibler information for each single candidate run. When computing the information divergence for combinations of different runs it is little extra work to also compute the variance inflation factor for each model both with and without a block factor. This gives the opportunity to take a possible block-effect into consideration.

The results from the examples considered indicate that there are a large number of follow-up designs which, depending on the models to discriminate, predictive density used and criterion to optimise, all are fairly good. What is important is to choose designs that are not especially bad.

References

Box, G. E. P. and Hill, W. J. (1967). Discrimination among Mechanistic Models, *Technometrics* **9**, 57–71.

Box, G. E. P. and Hunter, J. S. (1961). The 2^{k-p} Fractional Factorial Designs, *Technometrics* **3**, 311–351, 449–458.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). *Statistics for Experimenters*, John Wiley & Sons, New York.

Box, G. E. P. and Meyer, R. D. (1993). Finding the Active Factors in Fractionated Screening Experiments, *Journal of Quality Technology* **25**, 94–105.

Daniel, C. (1959). Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments, *Technometrics* **1**, 311–341.

Daniel, C. (1962). Sequences of Fractional Replicates in the 2^{p-q} Series, *Journal of the American Statistical Association* **57**, 403–429.

Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation*, John Wiley & Sons, New York.

Finney, D. J. (1945). The Fractional Replication of Factorial Arrangements, *Annals of Eugenics* **12**, 291–301.

Hamada, M. and Wu, C. F. J. (1992). Analysis of Designed Experiments with Complex Aliasing, *Journal of Quality Technology* **24**, 130–137.

Hynén, A. (1996). A Simple Method to Plan Follow-Up Runs to 12 Run Plackett and Burman Experiments, *Technical Report 4*, Division of Quality Technology and Management. Lindköping University, Sweden.

Keyes, T. K. and Levy, M. S. (1996). Goodness of Prediction Fit for Multivariate Linear Models, *Journal of the American Statistical Association* **91**, 191–197.

Kullback, S. (1968). *Information Theory and Statistics*, Dover Publications, Inc., New York.

Lin, D. K. J. and Draper, N. R. (1992). Projection Properties of Plackett and Burman Designs, *Technometrics* **34**, 423–428.

Marquardt, D. W. (1970). Generalized inverses, Ridge Regression, Biased Linear Estimation, and Non-Linear Estimation, *Technometrics* **12**, 591–612.

Meyer, R. D., Steinberg, D. M. and Box, G. E. P. (1996). Follow-Up Designs to Resolve Confounding in Multifactor Experiments (with discussion), *Technometrics* **38**, 303–332.

Murray, G. D. (1979). The Estimation of Multivariate Density Functions Using Incomplete Data, *Biometrika* **66**, 375–380.

Plackett, R. L. and Burman, J. P. (1946). The Design of Optimum Multifactorial Experiments, *Biometrica* **33**, 305–325.

Samset, O. (1998). Stepwise Regression under Factor Sparsity using a new Stopping Criterion, *Technical Report 12*, Department of Mathematical Sciences. The Norwegian University of Science and Technology, Norway.

Samset, O. (1999). Identifying Active Factors from Non-Geometric Plackett-Burman Designs and their Half-Fractions, *Technical Report 4*, Department of Mathematical Sciences. The Norwegian University of Science and Technology, Norway.

Tyssedal, J. and Samset, O. (1997). Analysis of the 12 Run Plackett-Burman Design, *Technical Report 8*, Department of Mathematical Sciences. The Norwegian University of Science and Technology, Norway.