Study of the Box-Meyer Method for Finding Active Factors in Screening Experiments

Oddgeir Samset and John Tyssedal
Department of Mathematical Sciences
The Norwegian University of Science and Technology
N-7034 Trondheim, Norway

February 12, 1998

Abstract

One of the suggested methods for identifying active factors in screening experiments is the Bayesian approach presented by Box and Meyer (1993). This method has been used successfully in the analysis of a number of both real and simulated data. To our experience however, there are situations where the method appears to be somewhat dependent of the size of both effects and error variance compared with a traditional regression analysis. In addition, when using the 12 run Plackett-Burman design the method often has problem with identifying more than three active factors. In this paper we look closer into the performance of the Box-Meyer method on a few simulated examples using both a 2_V^{5-1} fractional factorial design and a 12 run Plackett-Burman design. A simple modification of the Box-Meyer method, where non-significant terms in the individual models are removed, seems to reduce the described problems.

Key words: Factor sparsity; Plackett-Burman design; Regression; Variable selection.

1 Introduction

Box and Meyer (1993) present a new approach for identifying active factors in different types of screening experiments. Based on the principle of factor sparsity a set of models is stated. For every combination of potential active factors, a model here involves all main effects and interactions to a specified order. A Bayesian methodology is then used for estimating the posterior probability of each model given the observed response. Finally, the marginal posterior probabilities that an individual factor is active are computed by summing over the subset of models containing the factor.

This method is interesting by many reasons. First, the Box-Meyer method provides a way of analysing experiments from different types of screening designs such as 2^{k-p} fractional factorial designs (Finney, 1945; Box and Hunter, 1961) and non-geometric Plackett-Burman designs (Plackett and Burman, 1946). This method does not require orthogonal columns and can therefore handle situations where there are missing or additional data in the experiment. Secondly, the Bayesian methodology allows experimenters to include prior information into the analysis, and thereby to investigate how various prior assumptions may alter the results of the analysis. Thirdly, based on the probability calculations of this method Meyer, Steinberg and Box (1996) have also developed an approach for constructing follow-up experiments to resolve ambiguities from the initial design.

The Box-Meyer method has been used successfully in the analysis of a number of simulated examples from a variety of models. However, Tyssedal and Samset (1997) demonstrate that in some situations this method is somewhat dependent of the size of both effects and error variance compared with a traditional regression analysis. For instance, as the difference in size between the largest and smallest effects in the true model increases, the Box-Meyer method seems to have problem with identifying the smallest effects even if these are highly significant. Thus, in such situations there is a danger that active factors might be overlooked. Another problem we have experienced is difficulty with identifying more than three active factors when using the 12 run PB design. The same problem is also reported by Hynén (1996). However, the 12 run PB design has in many situations potential of identifying more than three active factors, see Tyssedal and Samset (1997) and Samset (1998).

In this paper we will look closer into the performance of the Box-Meyer method. This is done by focusing on the above mentioned problems in a few simulated examples using both a 2_V^{5-1} fractional factorial designs and a 12 run PB design. A short overview of the method as presented by Box and Meyer (1993) and Meyer and Box (1992) is given in Section 2. In Section 3 we illustrate the dependence of the size of effects using the 2_V^{5-1} fractional factorial design. The Box-Meyer method is here compared with both half-normal plot and best subsets regression. In Section 4 we look closer into the expression for the posterior probability for a model given the response, and simplify this by assuming that the model has an orthogonal matrix of regressor variables. This results in an expression including the well-known ordinary least squares estimate of the regression coefficients, and will therefore be helpful providing a better understanding of how the method performs. Based on these results we propose a minor modification to the Box-Meyer method in Section 5. The example considered in Section 3 is here reanalysed in view of the modification. In Section 6 we present some further examples with three and four active factors where the 12 run PB design is used for screening.

2 Methods for Analysing Screening Experiment

2.1 Alternative Methods

Non-replicated 2^{k-p} fractional factorial designs, which coincide with geometric PB designs, are usually analysed using normal and half-normal plot (Daniel, 1959; 1976) or Bayes-plot (Box and Meyer, 1986). For these designs all contrast columns corresponding to main effects and interactions are either orthogonal of fully confounded. One can therefore use such plots to identify the distinct contrasts that are too large to be explained by noise alone. However, if effects are fully confounded, many models that explain the data equally well can be specified, and may thereby make it difficult to identify the active factors. The common way to distinguish between such models is to perform a follow-up experiment to resolve ambiguities.

For non-geometric PB designs, contrast columns corresponding to main effects are orthogonal, but contrast columns corresponding to interactions can be both mutually correlated and correlated to main effect contrasts. As a result these designs have much more complex alias structures where effects are partially confounded. Until recently this has been regarded as a disadvantage in the analysis of experiments from these designs. However, as pointed out by Hamada and Wu (1991), designs with partially confounded effects have potential to simultaneously estimate both main effects and two-factor interactions under factor sparsity. Compared with 2^{k-p} fractional factorial designs one might therefore expect that non-geometric PB designs have a larger potential for separating out fewer or even a single model that give plausible explanations of the data. A traditional interpretation of the well-known normal and half-normal plot for non-geometric PB designs can be rather misleading in the presence of active two-factor interactions, resulting in active factors to appear as inert and vice versa. Thus, these designs seem to require a more thorough analysis of the experiments than the 2^{k-p} fractional factorials.

Hamada and Wu (1992) propose a procedure based on iterative use of forward selection. They start by using a traditional method such as half-normal plot to identify all main effect contrasts and two-factor interaction contrasts orthogonal to these that can not be explained by noise alone. Forward regression is then used to select the significant among these contrasts and two-factor interaction contrasts containing at least one on the factors already found. Main effects not initially entered into the model are reconsidered using forward regression together with terms already found. The two last steps are repeated until the model stabilises.

Chipman, Hamada and Wu (1997) apply a Bayesian approach using the SSVS (stochastic search variable selection) algorithm of George and McCulloch (1993) and Gibbs sampling. Their method gives opportunity to model relations, such as effects heredity, between main effects and terms of higher order by specifying hierarchical priors for interactions.

Tyssedal and Samset (1997) use an alternative interpretation of the half-normal plot where the alias-structure of the design is taken into account. It is demonstrated that such plots can give very valuable information in the presence of a few two-factor interactions.

Samset (1998) considers three variable selection techniques and compared them with the performance of best subsets regression. One of the methods is similar to best subsets regression but reduce the number of models to examine by guiding the inclusion of terms according to the principle of effects heredity. The other two methods are based on factor sparsity and the same models as proposed by Box and Meyer (1993). One method just regresses individual models to the response and presents the results as in best subsets regression. The other uses forward regression on the individual models. All these methods show promising results when they were used for reanalysing the simulated examples of Hamada and Wu (1992).

2.2 The Box-Meyer Method

In this section we give a brief outline of the method presented by Box and Meyer (1993) and Meyer and Box (1992).

Let $x_1, \ldots, x_k$ be k different input factors and y the response. Assume that there is a set of models $M_0, M_1, \ldots, M_m$ for which one of them hopefully can be used as a good approximation of the relation between the response and the input factors. The predictive density of $\boldsymbol{y} = (y_1, \ldots, y_n)'$ given the model M_i is then

$$f(\boldsymbol{y}|M_i) = \int_\Theta f(\boldsymbol{y}|M_i, \boldsymbol{\theta}_i) f(\boldsymbol{\theta}_i|M_i) d\boldsymbol{\theta}_i \tag{2.1}$$

where $f(\boldsymbol{y}|M_i, \boldsymbol{\theta}_i)$ describes the sampling distribution of $\boldsymbol{y}$ given the model and the parameters $\boldsymbol{\theta}_i$, and $f(\boldsymbol{\theta}_i|M_i)$ is the prior probability density of $\boldsymbol{\theta}_i$ given the model. The posterior probability of the model given the data is then found using Bayes theorem

$$p(M_i|\boldsymbol{y}) = \frac{p(M_i) f(\boldsymbol{y}|M_i)}{\sum_{j=0}^{m} p(M_j) f(\boldsymbol{y}|M_j)}. \tag{2.2}$$

Finally, the marginal posterior probability that factor j is active is found by summing over all models containing factor j

$$P_j = \sum_{M_i\,:\ \text{factor } j \text{ active}} p(M_i|\boldsymbol{y}). \tag{2.3}$$

In order to compute $p(M_i|\boldsymbol{y})$ the following assumptions are made:

1. The probability that any factor is active is π. The prior probability of model M_i is then $p(M_i) = \pi^{f_i}(1-\pi)^{k-f_i}$ where f_i is the number of active factors in this model.
2. The models are linear with independently and normally distributed noise. Thus, M_i is given by $\boldsymbol{y} = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}$ and the probability density $f(\boldsymbol{y}|M_i, \boldsymbol{\theta}_i)$ is

$$f(\boldsymbol{y}|M_i, \sigma, \boldsymbol{\beta}_i) \propto \sigma^{-n}\exp\{-(\boldsymbol{y}-\boldsymbol{X}_i\boldsymbol{\beta}_i)'(\boldsymbol{y}-\boldsymbol{X}_i\boldsymbol{\beta}_i)/2\sigma^2\}.$$

   Further $(\beta_{1i}, \ldots, \beta_{t_i i})'$ has a $N(\boldsymbol{0}, \gamma^2\sigma^2\boldsymbol{I})$ prior where t_i is the number of terms in the model additional to the intercept. The intercept β_{0i} and $\log(\sigma)$ have non-informative priors.

Box and Meyer then show that

$$p(M_i|\boldsymbol{y}) \approx C\left(\frac{\pi}{1-\pi}\right)^{f_i}\gamma^{-t_i}\frac{|\boldsymbol{X}_0'\boldsymbol{X}_0|^{1/2}}{|\Gamma_i + \boldsymbol{X}_i'\boldsymbol{X}_i|^{1/2}}\left(\frac{S(\hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i'\Gamma_i\hat{\boldsymbol{\beta}}_i}{S(\hat{\boldsymbol{\beta}}_0)}\right)^{-(n-1)/2} \tag{2.4}$$

where

$$\begin{aligned}
\Gamma_i &= \frac{1}{\gamma^2}\begin{pmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{t_i}\end{pmatrix} \\
\hat{\boldsymbol{\beta}}_i &= (\Gamma_i + \boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y} \\
S(\hat{\boldsymbol{\beta}}_i) &= (\boldsymbol{y}-\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)'(\boldsymbol{y}-\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)
\end{aligned}$$

Instead of using the full Bayesian specification where also γ is taken into account

$$p(M_i|\boldsymbol{y}) = \frac{\int_0^\infty [p(M_i|\boldsymbol{y},\gamma)/p(M_0|\boldsymbol{y},\gamma)]d\gamma}{\int_0^\infty [1/p(M_0|\boldsymbol{y},\gamma)]d\gamma}$$

the authors argue that $p(\gamma|\boldsymbol{y}) \propto 1/p(M_0|\boldsymbol{y})$. This means that the value of γ that minimises the probability of no active factors also maximises the posterior density $p(\gamma|\boldsymbol{y})$, see Meyer and Box (1992). Thus the procedure is to compute $p(M_0|\boldsymbol{y})$ for different values of γ and then choose the value of γ that minimises $p(M_0|\boldsymbol{y})$ or equivalently maximises $p(\gamma|\boldsymbol{y})$. Due to the unknown scaling constant in (2.4) this implies computing $p(M_i|\boldsymbol{y})$ for all models for each value of γ.

Based on the principle of factor sparsity it seems reasonable to first examine models with a low number of factors. However, since the number of active factors is unknown all possible combinations of one factor, two factors and so on have to be examined. For a given combination of potential active factors Box and Meyer choose to include all main effects and all interactions to a specified order. Thus if x_1, x_2 and x_3 are assumed to be the potential active factors, the model including main effect and two-factor interactions is given by

$$E(y) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3.$$

Use of such expanded models (submodels) for a given subset of potential active factors seems very attractive. This will substantially reduce the total number of models to examine compared with a more detailed all subsets strategy. One therefore has opportunity to perform a systematic search for the active factors within reasonable computer time.

Hynén (1996) recommends that the Box-Meyer method is carried out with varying model complexity. Thus, results from different choices of maximum number of active factors and maximum order of interactions to include.

3 Example 1

In this section we use a 2_V^{5-1} fractional factorial design to compare the performance of the Box-Meyer method with results from both half-normal plot and best subsets regression.

The responses are simulated from the model $y = ax_1 + x_2 + 0.5x_1x_3 + \epsilon$ where $\epsilon \sim N(0, c^2)$ using the 2_V^{5-1} fractional factorial design with generator $I = ABCDE$. Both a and c will be varied. For better to compare the effect of varying a and c all computations are based on the same noise vector that is a realization from a standard normal distribution $N(0,1)$

$$\epsilon_s = (1.17\ 0.63\ 0.08\ 0.35\ \text{-}0.70\ 1.70\ 0.06\ 1.80\ 0.26\ 0.87\ \text{-}1.45\ \text{-}0.70\ 1.25\ \text{-}0.64\ 0.58\ \text{-}0.36)'.$$

The appropriate ϵ is then found in each situation by multiplying ϵ_s with the specific c. In all computations we assume three-factor and higher order interactions to be negligible. Thus, since we are using a design of resolution V matrices of regressor variables (including constant column, main effects and two-factor interactions) are orthogonal for every submodel used in the Box-Meyer method. Up to three factors are included into the models and we use $\pi = 0.25$.

Figure 1 shows half-normal plot, marginal posterior probabilities and probability density of gamma given the data (scaled to 1 at the mode) for the response described above. In the three situations considered we have used $a = 1$ (leftmost), $a = 3$ (centre) and $a = 5$ (rightmost). In all situations $c = 0.3$. All half-normal plots strongly support that the terms x_1, x_2 and x_1x_3 are active. Since the realization of the noise is the same, the only difference between the three situations is the size of the main effect of x_1. Hence the distances between the size of the two-factor interaction x_1x_3 and the inert terms are the same in all situations. This is however somewhat obscured in the figure due to different scaling of axes. From the marginal posterior probabilities we see that as the value of a increases the probability of factor x_3 is reduced dramatically. In all of the situations considered all submodels except those containing the factors (x_1), (x_1, x_2), (x_1, x_2, x_3) have a probability $p(M_i|\boldsymbol{y})$ close to zero. Thus, as a increases the model with factors (x_1, x_2, x_3) seems to get less important compared with the two other models.

In Figure 2 results from best subsets regression are presented for $a = 1$ (leftmost) and $a = 5$ (rightmost). The four 'best' models according to residual mean square error are listed for models with zero to three terms (in addition to the intercept). Also best subsets regression indicates the model with the terms x_1, x_2 and x_1x_3 as the most plausible. A closer look at this model shows that β_{13} is significant at 0.0% level in both situations. This reveals that the Box-Meyer method here is somewhat more sensitive to the size of effects and thereby the choice of levels compared with both half-normal plot and best subsets regression.

In Figure 3 we have given the results for $c = 0.2$ (leftmost), $c = 0.4$ (centre) and $c = 0.6$ (rightmost). In all situations $a = 3$. As the standard deviation of the noise increases it is obvious that it becomes more difficult to detect small signals. This is also the impression we get from both half-normal plots and marginal probabilities. The results from best subsets regression for $a = 1$ (leftmost) and $a = 5$ (rightmost) are given in Figure 4. For the model containing the terms x_1, x_2 and x_1x_3 the parameter β_{13} is however significant at 0.0% and 0.5% levels for $c = 0.2$ and $c = 0.6$, respectively. As a comparison, for the model containing the terms x_1, x_2 and x_2x_5 the parameter β_{25} is significant at 48.7% and 10.7%, respectively. Thus, in these situations the Box-Meyer method also seems to be more sensitive to an increase in c compared with half-normal plot and best subsets regression.

In Figure 1 and Figure 3 only main effects and two-factor interactions were included into the submodels for the Box-Meyer method. If also three-factor interactions were considered the marginal probability of factor x_3 would be even less. The marginal probability that factor x_3 is active is then 0.68, 0.07 and 0.01 for the three situations in Figure 1, and 0.65, 0.02 and 0.01 for the three situations in Figure 3, respectively.

4 The Posterior Probability of a Model

Based on the results from the previous section there are some unanswered questions related to the Box-Meyer method and how it performs. This especially concerns how weighting is achieved between models with same number of terms and between models with different number of terms. This also includes estimation of gamma and how this estimate will act upon the different types of weighting.

4.1 Weighting Due to Factors, Terms and Model Fit

Reconsider the posterior probability of a model given the data in (2.4)

$$p(M_i|\boldsymbol{y}) \approx C\left(\frac{\pi}{1-\pi}\right)^{f_i} \gamma^{-t_i} \frac{|\boldsymbol{X}_0'\boldsymbol{X}_0|^{1/2}}{|\Gamma_i + \boldsymbol{X}_i'\boldsymbol{X}_i|^{1/2}} \left(\frac{S(\hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i'\Gamma_i\hat{\boldsymbol{\beta}}_i}{S(\hat{\boldsymbol{\beta}}_0)}\right)^{-(n-1)/2}$$

Box and Meyer (1993) point out that

$$\left(\frac{\pi}{1-\pi}\right)^{f_i} \gamma^{-t_i} \frac{|\boldsymbol{X}_0'\boldsymbol{X}_0|^{1/2}}{|\Gamma_i + \boldsymbol{X}_i'\boldsymbol{X}_i|^{1/2}} \tag{4.1}$$

is a penalty for increasing the number of variables into the model,while

$$\left(\frac{S(\hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i'\Gamma_i\hat{\boldsymbol{\beta}}_i}{S(\hat{\boldsymbol{\beta}}_0)}\right)^{-(n-1)/2} \tag{4.2}$$

is a measure of how well the model fits.

Regarding (4.1) it is important to notice that since Box and Meyer use submodels where all main effects and interaction to a specified order are included, the number of terms t_i is directly given from the number of factors f_i. However, not all terms in a submodel will usually be significant. It seems therefore somewhat hard to punish a model based on the number of terms when not considering whether they are significant of not. If we define

$$\begin{aligned} w_1 &= \left(\frac{\pi}{1-\pi}\right)^{f_i} \\ w_2 &= \gamma^{-t_i} \frac{|\boldsymbol{X}_0'\boldsymbol{X}_0|^{1/2}}{|\Gamma_i + \boldsymbol{X}_i'\boldsymbol{X}_i|^{1/2}} \end{aligned}$$

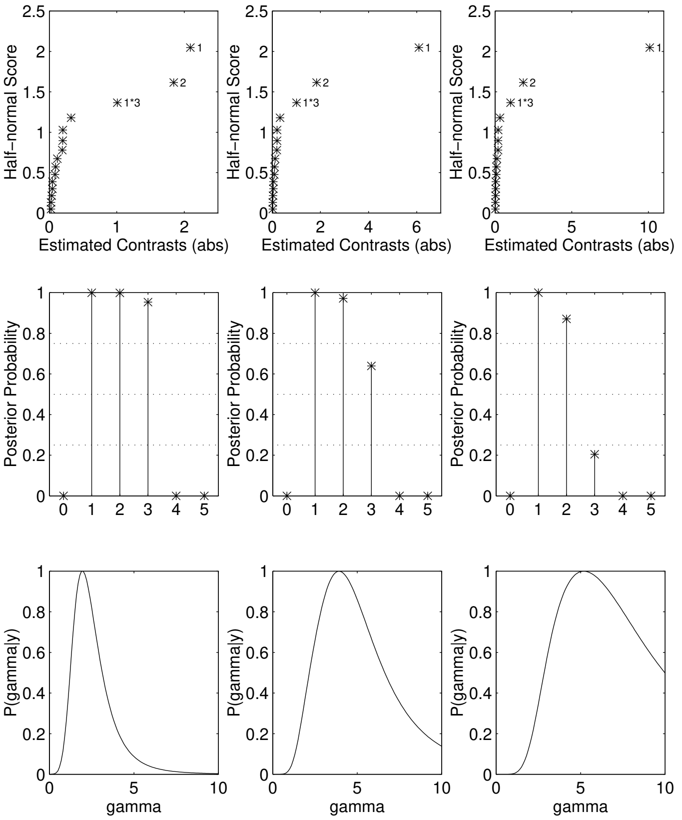

Figure 1: Half-normal plot, marginal posterior probabilities and (mode-scaled) probability density of γ. Response is given by $y = ax_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $a = 1, 3, 5$ and $\epsilon \sim N(0, 0.3^2)$ using the 2_V^{5-1} design.

```
 Number of runs in design      :   16
 Number of factors in design   :    5
 Max number of terms           :    3
 Number of models examined     :  576
 "Best model" criteria         : Residual MSe

Terms   R_sq  R_sqa     s    Terms in model
-------------------------------------------
    0   0.00   0.00  1.55
    1  48.32  44.63  1.16    1
    1  37.59  33.13  1.27    2
    1  11.25   4.91  1.51    1*3
    1   1.15   0.00  1.60    2*5
    2  85.91  83.74  0.63    1 2
    2  59.57  53.35  1.06    1 1*3
    2  49.47  41.70  1.19    1 2*5
    2  48.83  40.96  1.19    2 1*3
    3  97.16  96.45  0.29    1 2 1*3
    3  87.06  83.83  0.62    1 2 2*5
    3  86.36  82.94  0.64    1 2 1*4
    3  86.35  82.93  0.64    1 2 4
```

```
 Number of runs in design      :   16
 Number of factors in design   :    5
 Max number of terms           :    3
 Number of models examined     :  576
 "Best model" criteria         : Residual MSe

Terms   R_sq  R_sqa     s    Terms in model
-------------------------------------------
    0   0.00   0.00  5.33
    1  95.61  95.30  1.16    1
    1   3.19   0.00  5.43    2
    1   0.95   0.00  5.49    1*3
    1   0.10   0.00  5.51    2*5
    2  98.80  98.62  0.63    1 2
    2  96.57  96.04  1.06    1 1*3
    2  95.71  95.05  1.19    1 2*5
    2  95.65  94.98  1.19    1 1*4
    3  99.76  99.70  0.29    1 2 1*3
    3  98.90  98.63  0.62    1 2 2*5
    3  98.84  98.55  0.64    1 2 1*4
    3  98.84  98.55  0.64    1 2 4
```

Figure 2: Best subsets regression for $y = 1x_1 + x_2 + 0.5x_1x_3 + \epsilon$ and $y = 5x_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $\epsilon \sim N(0, 0.3^2)$ using the 2_V^{5-1} design.

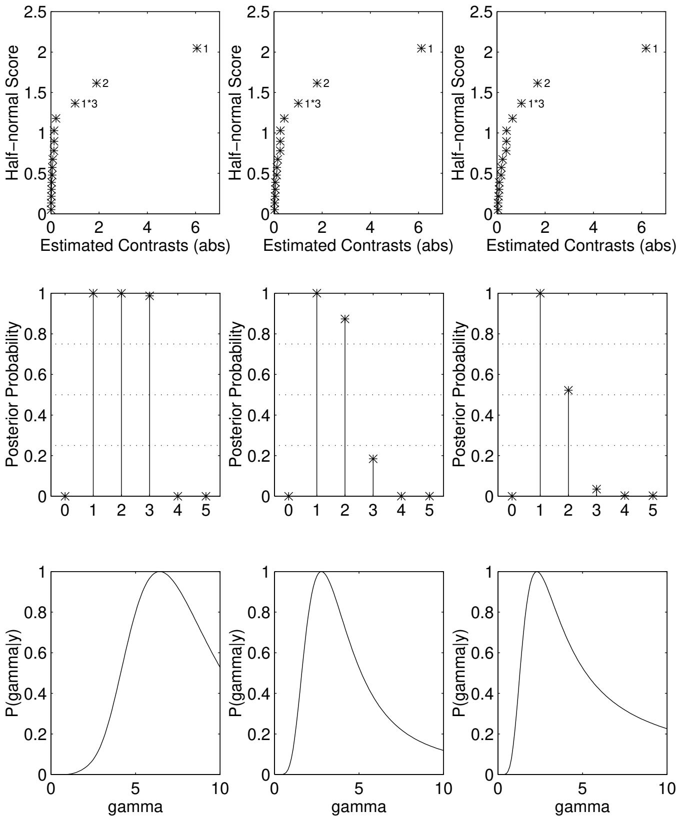

Figure 3: Half-normal plot, marginal posterior probabilities and (mode-scaled) probability density of γ. Response is given by $y = 3x_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $\epsilon \sim N(0, c^2)$ and $c = 0.2, 0.4, 0.6$ using the 2_V^{5-1} design.

```
 Number of runs in design     :   16
 Number of factors in design  :    5
 Max number of terms          :    3
 Number of models examined    :  576
 "Best model" criteria        : Residual MSe

Terms   R_sq  R_sqa     s    Terms in model
-------------------------------------------
    0   0.00   0.00  3.32
    1  88.62  87.80  1.16    1
    1   8.67   2.15  3.29    2
    1   2.44   0.00  3.40    1*3
    1   0.11   0.00  3.44    2*5
    2  97.28  96.87  0.59    1 2
    2  91.05  89.68  1.07    1 1*3
    2  88.73  86.99  1.20    1 2*5
    2  88.66  86.91  1.20    1 1*4
    3  99.72  99.66  0.20    1 2 1*3
    3  97.40  96.75  0.60    1 2 2*5
    3  97.33  96.66  0.61    1 2 1*4
    3  97.33  96.66  0.61    1 2 4
```

```
 Number of runs in design     :   16
 Number of factors in design  :    5
 Max number of terms          :    3
 Number of models examined    :  576
 "Best model" criteria        : Residual MSe

Terms   R_sq  R_sqa     s    Terms in model
-------------------------------------------
    0   0.00   0.00  3.39
    1  88.62  87.81  1.18    1
    1   6.60   0.00  3.39    2
    1   2.40   0.00  3.47    1*3
    1   0.97   0.00  3.49    2*5
    2  95.22  94.48  0.80    1 2
    2  91.02  89.63  1.09    1 1*3
    2  89.58  87.98  1.18    1 2*5
    2  88.99  87.30  1.21    1 1*4
    3  97.62  97.02  0.59    1 2 1*3
    3  96.18  95.23  0.74    1 2 2*5
    3  95.59  94.49  0.80    1 2 1*4
    3  95.58  94.48  0.80    1 2 4
```

Figure 4: Best subsets regression for $y = 3x_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $\epsilon \sim N(0, 0.2^2)$ and $N(0, 0.6^2)$ using the 2_V^{5-1} design.

then w_1 is a penalty for increasing the number of factors and w_2 is the penalty for the number of terms in addition to the intercept in the given model. Thus, w_1 might be viewed as penalty due to factor sparsity, and w_2 as penalty due to effect sparsity within a model with a given number of factors.

Since (4.2) is a measure of how well the model fits, it is interesting to compare this measure with the residual mean square error from ordinary least squares. Now $S(\hat{\boldsymbol{\beta}}_i)$ can be written

$$S(\hat{\boldsymbol{\beta}}_i) = S(\hat{\boldsymbol{b}}_i) + (\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{b}}_i)'\boldsymbol{X}_i'\boldsymbol{X}_i(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{b}}_i)$$

where $\hat{\boldsymbol{b}}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y}$ is the ordinary least squares estimate of $\boldsymbol{\beta}_i$, and where $S(\hat{\boldsymbol{b}}_i) = (\boldsymbol{y} - \boldsymbol{X}_i\hat{\boldsymbol{b}}_i)'(\boldsymbol{y} - \boldsymbol{X}_i\hat{\boldsymbol{b}}_i)$. We can therefore express (4.2) as

$$\left(\frac{S(\hat{\boldsymbol{b}}_i) + (\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{b}}_i)'\boldsymbol{X}_i'\boldsymbol{X}_i(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{b}}_i) + \hat{\boldsymbol{\beta}}_i'\Gamma_i\hat{\boldsymbol{\beta}}_i}{S(\hat{\boldsymbol{\beta}}_0)}\right)^{-(n-1)/2}. \tag{4.3}$$

In order to compare the ordinary least squares fit with the contribution from the Bayesian model it seems natural to define

$$w_3 = \left(\frac{w_{31} + w_{32}}{S(\hat{\boldsymbol{\beta}}_0)}\right)^{-(n-1)/2}$$

where

$$\begin{aligned} w_{31} &= S(\hat{\boldsymbol{b}}_i) \\ w_{32} &= (\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{b}}_i)'\boldsymbol{X}_i'\boldsymbol{X}_i(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{b}}_i) + \hat{\boldsymbol{\beta}}_i'\Gamma_i\hat{\boldsymbol{\beta}}_i \end{aligned}$$

4.2 Orthogonal Matrix of Regressor Variables

Assume that all submodels have orthogonal matrices of regressor variables, i.e. $\boldsymbol{X}_i'\boldsymbol{X}_i = n\boldsymbol{I}$. Then it is easy to verify that

$$\begin{aligned} \gamma^{-t_i}|\Gamma_i + \boldsymbol{X}_i'\boldsymbol{X}_i|^{-1/2} &= n^{-1/2}(n\gamma^2+1)^{-t_i/2} \\ (\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{b}}_i)'\boldsymbol{X}_i'\boldsymbol{X}_i(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{b}}_i) &= n(n\gamma^2+1)^{-2}\hat{\boldsymbol{b}}_i^{*'}\hat{\boldsymbol{b}}_i^* \\ \hat{\boldsymbol{\beta}}_i'\Gamma_i\hat{\boldsymbol{\beta}}_i &= n^2\gamma^2(n\gamma^2+1)^{-2}\hat{\boldsymbol{b}}_i^{*'}\hat{\boldsymbol{b}}_i^* \end{aligned}$$

where $\hat{\boldsymbol{b}}_i$ is the usual least squares estimate of $\boldsymbol{\beta}_i$, and $\hat{\boldsymbol{b}}_i^*$ is equal to $\hat{\boldsymbol{b}}_i$ but with the intercept term put to zero, i.e. $\hat{\boldsymbol{b}}_i^* = (0, \hat{b}_{1i}, \ldots, \hat{b}_{t_i i})'$. Thus, for an orthogonal matrix of regressor variables we can simplify the expression for the posterior probability to

$$p(M_i|\boldsymbol{y}) \approx C\left(\frac{\pi}{1-\pi}\right)^{f_i}(n\gamma^2+1)^{-t_i/2}\left(\frac{S(\hat{\boldsymbol{b}}_i) + n(n\gamma^2+1)^{-1}\hat{\boldsymbol{b}}_i^{*'}\hat{\boldsymbol{b}}_i^*}{S(\hat{\boldsymbol{b}}_0)}\right)^{-(n-1)/2}. \tag{4.4}$$

The weights introduced above then simplify to

$$\begin{aligned} w_1 &= \left(\frac{\pi}{1-\pi}\right)^{f_i} \\ w_2 &= (n\gamma^2+1)^{-t_i/2} \\ w_3 &= \left(\frac{w_{31}+w_{32}}{S(\hat{\boldsymbol{b}}_0)}\right)^{-(n-1)/2} = \left(\frac{S(\hat{\boldsymbol{b}}_i) + n(n\gamma^2+1)^{-1}\hat{\boldsymbol{b}}_i^{*'}\hat{\boldsymbol{b}}_i^*}{S(\hat{\boldsymbol{b}}_0)}\right)^{-(n-1)/2} \end{aligned}$$

Models with same number of potential active factors are entirely compared according to w_3. For different number of potential active factors models are compared based on all three weights w_1, w_2 and w_3.

We know from the description of the model that $(\beta_{1i}, \ldots, \beta_{t_i i})'$ has a $N(\mathbf{0}, \gamma^2\sigma^2\boldsymbol{I})$ prior. Thus, the size of the individual parameters will in some way direct the value of γ. However, it seems reasonable to assume that w_{32} not should be dominant over w_{31}, i.e. that the residual mean square error from ordinary least squares is not dominated by the length of the least square estimate of $\boldsymbol{\beta}$. Consider the three situations illustrated in Figure 1. Here w_{32} increases due to the increase in a while w_{31} is constant for a model with a given number of factors. As the size of the effect and thereby the length of $\boldsymbol{\beta}$ increases, the value of γ should also increase to reduce the influence of w_{32}. The increase in γ appears to be consistent with the situations in Figure 1. However, as γ increases, w_2 will decrease and thus put much less weight on models with large values of t_i. This might explain why the model with factors (x_1, x_2, x_3) is getting less important as a increases.

A value of γ must be found that gives a compromise between the weights. The number of terms t_i in the model appears to be an important parameter with respect to this compromise.

4.3 Revisiting Example 1

We here reconsider example 1 where we used the 2_V^{5-1} fractional factorial design, and look at the influence of the respective weights given above. Since the design is of resolution V and we neglect three-factor and higher order interactions, matrices of regressor variables for the submodels are orthogonal, and we can use the simplified weights.

The results for the different situations in Figure 1 and Figure 3 are summarised in Table 1 and Table 2. The most interesting results for these situations are:

1. For the estimated values of γ in these situations the punishment due to the number of terms w_2 may be considerable larger than for the number of factors w_1. This is especially true for large values of f_i. Hence the punishment due to the number of factors often seems to go mainly through w_2 instead of w_1. Eliminating the effect of w_1 by setting $\pi = 0.5$ improve the marginal probabilities only slightly and the problem is still obvious.

2. As the value of a increases the increase in the estimated value of γ does not prevent w_{32} from being of the same size or even larger than w_{31}. This is especially true for large values of t_i. Thus, the size of the effects may have considerably influence on the measure of fit w_3.

5 A Modification of the Box-Meyer Method

The results from the previous sections indicate that for models with a large number of potential active factors the value of t_i is too high. This results in a very low weight on these models. Hence, in addition to factor sparsity one should also consider effect sparsity in individual models. One possible modification to achieve this is to remove all non-significant terms, i.e. terms with small standardised coefficients, from each model.

From earlier we have $\hat{\boldsymbol{\beta}}_i = (\Gamma_i + \boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y}$ and $S(\hat{\boldsymbol{\beta}}_i) = (\boldsymbol{y} - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)'(\boldsymbol{y} - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_i)$. Then the variance of $\hat{\boldsymbol{\beta}}_i$ is given by

$$Var(\hat{\boldsymbol{\beta}}_i) = (\Gamma_i + \boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{X}_i(\Gamma_i + \boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\sigma^2 = \boldsymbol{V}\sigma^2$$

Table 1: Summary for response $y = ax_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $\epsilon \sim N(0, 0.3^2)$.

a	Factors	$\hat{\gamma}$	$p(M_i\|\boldsymbol{y})$	w_1	w_2	w_{31}	w_{32}	w_3
1	1	1.95	0.001	3.33 e-1	1.27 e-1	1.87 e1	2.83 e-1	1.26 e02
	1 2		0.045	1.11 e-1	2.06 e-3	5.09	5.02 e-1	1.21 e06
	1 2 3		0.954	3.70 e-2	4.23 e-6	8.37 e-1	5.71 e-1	3.73 e10
3	1	3.92	0.028		6.37 e-2		6.01 e-1	1.08 e07
	1 2		0.333		2.58 e-4		6.56 e-1	9.52 e10
	1 2 3		0.639		6.65 e-8		6.73 e-1	2.13 e15
5	1	5.19	0.128		4.81 e-2		9.43 e-1	1.06 e10
	1 2		0.667		1.11 e-4		9.74 e-1	7.11 e13
	1 2 3		0.204		1.24 e-8		9.84 e-1	5.86 e17

Table 2: Summary for response $y = 3x_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $\epsilon \sim N(0, c^2)$.

c	Factors	$\hat{\gamma}$	$p(M_i\|\boldsymbol{y})$	w_1	w_2	w_{31}	w_{32}	w_3
0.2	1	6.43	0.001	3.33 e-1	3.89 e-2	1.89 e1	2.22 e-1	1.10 e07
	1 2		0.012	1.11 e-1	5.86 e-5	4.50	2.43 e-1	3.78 e11
	1 2 3		0.987	3.70 e-2	3.44 e-9	3.72 e-1	2.50 e-1	1.57 e18
0.3	1	3.92	0.028		6.37 e-2	1.87 e1	6.01 e-1	1.08 e07
	1 2		0.333		2.58 e-4	5.09	6.56 e-1	9.52 e10
	1 2 3		0.639		6.65 e-8	8.37 e-1	6.73 e-1	2.13 e15
0.4	1	2.77	0.124		8.99 e-2	1.88 e1	1.21	8.91 e06
	1 2		0.690		7.26 e-4	5.90	1.31	1.84 e10
	1 2 3		0.183		5.28 e-7	1.49	1.35	2.01 e13
0.6	1	2.30	0.464		1.08 e-1	1.96 e1	1.78	6.24 e6
	1 2		0.495		1.26 e-3	8.21	1.92	1.71 e9
	1 2 3		0.026		1.59 e-6	3.35	1.97	2.13 e11

where σ^2 can be estimated by $\hat{\sigma}_i^2 = S(\hat{\boldsymbol{\beta}}_i)/(n - t_i - 1)$.

A relatively simple modification of the Box-Meyer method that only reduce the computing effort slightly is then as follows:

1. For each model compute $\hat{\boldsymbol{\beta}}_i$ and $S(\hat{\boldsymbol{\beta}}_i)$ as before.

2. Remove all terms in $\boldsymbol{X}_i$ where the corresponding $|\hat{\beta_{ij}}| < \delta V_{jj}^{1/2}\hat{\sigma}_i$. Here we have used $\delta = 2$ if not explicitly stated elsewhere, but one should preferable try different choices such as $\delta = 0, 1, 2$ to examine the influence of non-significant terms on the result. Using the value $\delta = 0$ prevents the modification from removing any terms, and thus coincides with the original Box-Meyer method.

3. Adjust t_i and Γ_i, repeat the computation of $\hat{\boldsymbol{\beta}}_i$ and $S(\hat{\boldsymbol{\beta}}_i)$ and proceed as before.

It should be noted that f_i is not updated accordingly in the implemented version of this modification. Let us assume that factors x_1, x_2, x_3 and x_4 are in the model. None of the terms containing factor x_4 are however significant and are excluded in the computation of

$p(M_i|\boldsymbol{y})$. In this situation we still use $f = 4$, and $p(M_i|\boldsymbol{y})$ for this model is also included when computing the marginal posterior probability that factor x_4 is active. If all terms containing factor x_4 are removed from the model this indicates that the three-factor model fits the data reasonable well. However, this model is already taken care of when considering all three-factor models, and should therefore not be receiving a high weight from the reduced four-factor model as well. One other possibility would be to first find all reduced models, remove any duplicate models and then find their probability. This would however increase the computing time.

The removal of terms may also be done iteratively by first removing the less significant term, repeat the computation of $\hat{\boldsymbol{\beta}}_i$ and $S(\hat{\boldsymbol{\beta}}_i)$, new removal and so on until no terms are removed. One may also consider expanding the model space by for example using all possible combinations of terms, i.e. all models from a best subsets regression, see the discussion by Chipman and Hamada in Meyer et al. (1996). However, such methods will increase the computing time significantly.

5.1 Revisiting Example 1 Using the Modified Box-Meyer Method

Figure 5 and Figure 6 show the result for the 2_V^{5-1} fractional factorial design considered earlier using the modification of the Box-Meyer method. Comparing these two figures with Figure 1 and Figure 3 show a significant change. This is especially true for Figure 5 where a is varied. Increasing a has now very little effect on the marginal posterior probabilities. Even an extreme situation as $a = 100$ and $c = 0.3$ still gives a marginal posterior probability of 0.90 for factor x_3. Figure 6 shows that the modified method still is dependent of the size of the error variance as it should be, but now to a less degree than before. Choosing $a = 3.0$ and $c = 0.6$, where β_{13} is significant at 0.5% level, gives a marginal posterior probability of 0.53 for factor x_3. Choosing $a = 3.0$ and $c = 0.8$, where β_{13} is significant at 2.2% level, gives a marginal posterior probability of 0.31 for factor x_3.

In Table 3 and Table 4 we have listed the summaries for the modified method as a comparison with the results in Table 1 and Table 2. An extra column is however added showing the significant terms in the respective models. Most striking is that the value of γ now is considerably larger than before. The weight w_1 is the same as before. Also w_{31} is approximately the same as before for the models listed showing that the non-significant terms in this situation give relatively little contribution to the ordinary least squares fit. Non-significant terms may however give larger contribution for the other models. The weight w_2 is considerably increased for the three-factor models due to a reduced value of t_i . For the situation where $c = 0.3$ a variation in a now result in an almost constant w_{32} which now is relatively small compared with w_{31}. Thus the length of $\hat{\boldsymbol{\beta}}$ has less influence than before. When varying c for $a = 3$ the weight w_{32} is still increasing with increasing value of c, but w_{32} is now smaller relatively to w_{31}. However, the major difference in this situation seems to be the reduction in w_2.

Removing the non-significant terms and thereby reducing t_i allows γ to increase. A larger value of γ gives higher weight on the residual mean square error from the ordinary least squares relative to the contribution from the length of the vector of estimated parameters. Thus the modified method gets less dependent of the size of the individual effects.

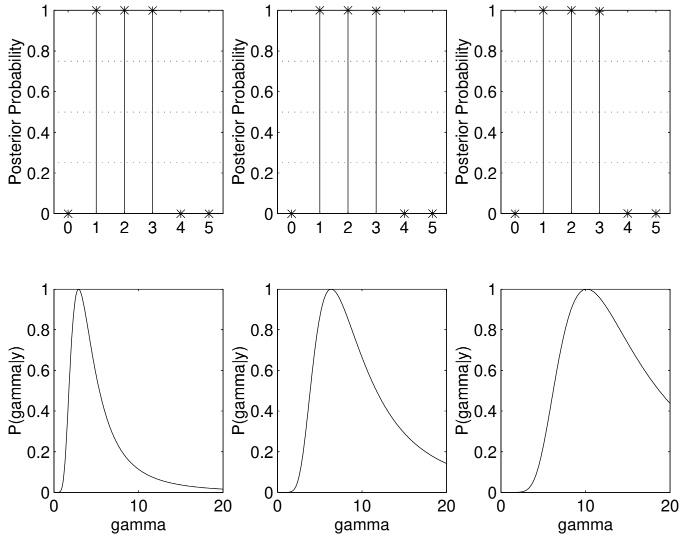

Figure 5: Marginal posterior probabilities and (mode-scaled) probability density of γ for the modified Box-Meyer method. Response is given by $y = ax_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $a = 1, 3, 5$ and $\epsilon \sim N(0, 0.30^2)$ using the 2_V^{5-1} design.

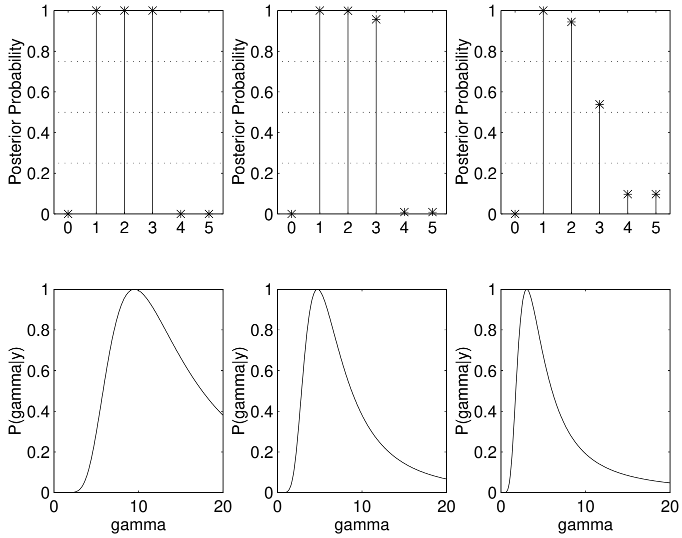

Figure 6: Half-normal plot, marginal posterior probabilities and (mode-scaled) probability density of γ for the modified Box-Meyer method. Response is given by $y = 3x_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $\epsilon \sim N(0, c^2)$ and $c = 0.2, 0.4, 0.6$ using the 2_V^{5-1} design.

Table 3: Summary for response $y = ax_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $\epsilon \sim N(0, 0.3^2)$ using the modified Box-Meyer method.

a	Factors	Terms	$\hat{\gamma}$	$p(M_i\|\boldsymbol{y})$	w_1	w_2	w_{31}	w_{32}	w_3
1	1	1	2.91	0.000	3.33 e-1	8.56 e-2	1.87 e1	1.28 e-1	1.34 e02
	1 2	1 2		0.001	1.11 e-1	7.33 e-3	5.09	2.28 e-1	1.74 e06
	1 2 3	1 2 1*3		0.999	3.70 e-2	6.27 e-4	1.03	2.57 e-1	7.41 e10
3	1	1	6.35	0.000		3.93 e-2		2.30 e-1	1.25 e07
	1 2	1 2		0.002		1.55 e-3		2.51 e-1	1.63 e11
	1 2 3	1 2 1*3		0.997		6.09 e-5		2.57 e-1	7.17 e15
5	1	1	10.15	0.000		2.46 e-2		2.47 e-1	1.38 e10
	1 2	1 2		0.003		6.06 e-4		2.55 e-1	1.81 e14
	1 2 3	1 2 1*3		0.995		1.49 e-5		2.58 e-1	7.99 e18

Table 4: Summary for response $y = 3x_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $\epsilon \sim N(0, c^2)$ using the modified Box-Meyer method.

c	Factors	Terms	$\hat{\gamma}$	$p(M_i\|\boldsymbol{y})$	w_1	w_2	w_{31}	w_{32}	w_3
0.2	1	1	9.51	0.000	3.33 e-1	2.63 e-2	1.89 e1	1.01 e-1	1.15 e07
	1 2	1 2		0.000	1.11 e-1	6.91 e-4	4.50	1.11 e-1	4.64 e11
	1 2 3	1 2 1*3		1.000	3.70 e-2	1.82 e-5	4.57 e-1	1.14 e-1	2.97 e18
0.3	1	1	6.35	0.000		3.93 e-2	1.87 e1	2.30 e-1	1.25 e07
	1 2	1 2		0.002		1.55 e-3	5.09	2.51 e-1	1.63 e11
	1 2 3	1 2 1*3		0.997		6.09 e-5	1.03	2.57 e-1	7.17 e15
0.4	1	1	4.73	0.000		5.28 e-2	1.88 e1	4.17 e-1	1.21 e07
	1 2	1 2		0.025		2.79 e-3	5.92	4.53 e-1	4.68 e10
	1 2 3	1 2 1*3		0.957		1.47 e-4	1.83	4.65 e-1	1.00 e14
0.6	1	1	3.05	0.024		8.17 e-2	1.96 e1	1.02	8.19 e06
	1 2	1 2		0.251		6.67 e-3	8.24	1.10	3.14 e09
	1 2 3	1 2 1*3		0.526		5.45 e-4	4.11	1.12	2.42 e11

6 Further Examples

6.1 Example 2

We will again reconsider the model from example 1 but now using the saturated 12 run Plackett-Burman design instead of the 2_V^{5-1} fractional factorial design. The 12 run Plackett-Burman design is given in Table 5. Thus, the response is still simulated from the model $y = ax_1 + x_2 + 0.5x_1x_3 + \epsilon$ where $\epsilon \sim N(0, c^2)$. Again we vary both a and c using the same values as before. The realization of the noise vector is here given by

$$\epsilon_s = (1.23\ 0.99\ \text{-}0.80\ \text{-}0.59\ 1.69\ 0.95\ \text{-}1.93\ 0.51\ 0.39\ \text{-}0.91\ \text{-}1.27\ 0.35)'.$$

Figure 7 shows half-normal plot, marginal probabilities and (mode-scaled) probability density of gamma for $a = 1$ (leftmost), $a = 3$ (centre) and $a = 5$ (rightmost) when $c = 0.3$. Here the half-normal plots, which only include the main effect contrasts, indicate that the contrasts corresponding to factor x_1 and x_2 are active. However, there are only weak indications that an additional factor is active. Tyssedal and Samset (1997) suggest that one also should consider generating additional half-normal plots of all two-factor interactions of the individual factors. Such plots are not included, but they reveal interesting information in these situations. The two active main effects corresponding to factors x_1 and x_1 are not appropriate to explain their patterns, and indicate an additional active interaction.

For the marginal posterior probabilities we experience the same dependency of a as for the 2_V^{5-1} design, now even more striking than before. The parameter β_{13} is here significant at 0.1% level in all three situations. Figure 8 shows the marginal posterior probabilities using the modified version of the Box-Meyer method. For the three situations given here all three factors now stand out as active. However, there still seem to be some dependency of a left. Note however that using $\delta = 2$ will here only remove x_2x_3 of the two inert terms in the model containing factors x_1, x_2 and x_3.

Figure 9 shows half-normal plot, marginal probabilities and (mode-scaled) probability density of gamma for $c = 0.2$ (leftmost), $c = 0.4$ (centre) and $c = 0.6$ (rightmost) when $a = 3$. The parameter β_{13} is here significant at 0.0%, 0.5% and 2.9% level for the three situations, respectively. The marginal posterior probabilities using the modified version are given in Figure 10. Also here $\delta = 2$ only removes the inert term x_2x_3.

Table 5: The 12 run Plackett-Burman design.

run	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
1	+	-	+	-	-	-	+	+	+	-	+
2	+	+	-	+	-	-	-	+	+	+	-
3	-	+	+	-	+	-	-	-	+	+	+
4	+	-	+	+	-	+	-	-	-	+	+
5	+	+	-	+	+	-	+	-	-	-	+
6	+	+	+	-	+	+	-	+	-	-	-
7	-	+	+	+	-	+	+	-	+	-	-
8	-	-	+	+	+	-	+	+	-	+	-
9	-	-	-	+	+	+	-	+	+	-	+
10	+	-	-	-	+	+	+	-	+	+	-
11	-	+	-	-	-	+	+	+	-	+	+
12	-	-	-	-	-	-	-	-	-	-	-

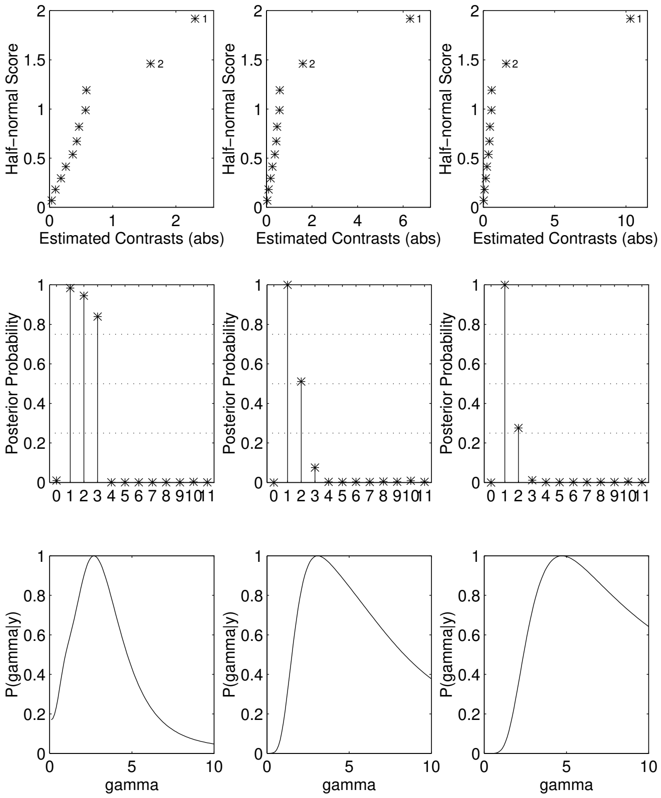

Figure 7: Half-normal plot, marginal posterior probabilities and (mode-scaled) probability density of γ. Response is given by $y = ax_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $a = 1, 3, 5$ and $\epsilon \sim N(0, 0.3^2)$ using the 12 run Plackett-Burman design.

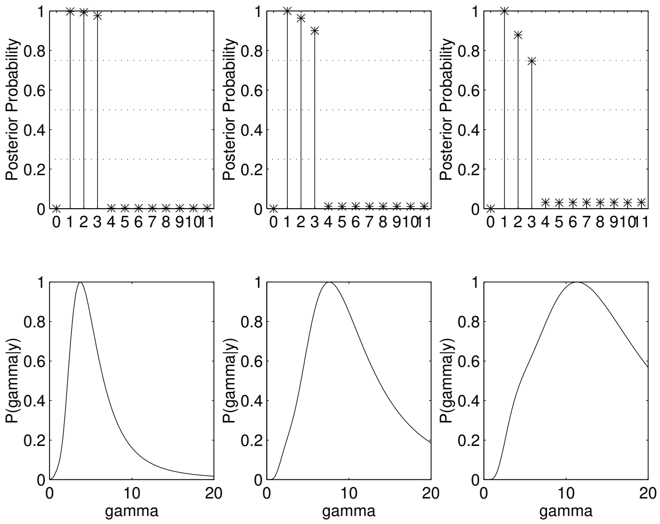

Figure 8: Marginal posterior probabilities and (mode-scaled) probability density of γ for the modified Box-Meyer method. Response is given by $y = ax_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $a = 1, 3, 5$ and $\epsilon \sim N(0, 0.3^2)$ using the 12 run Plackett-Burman design.

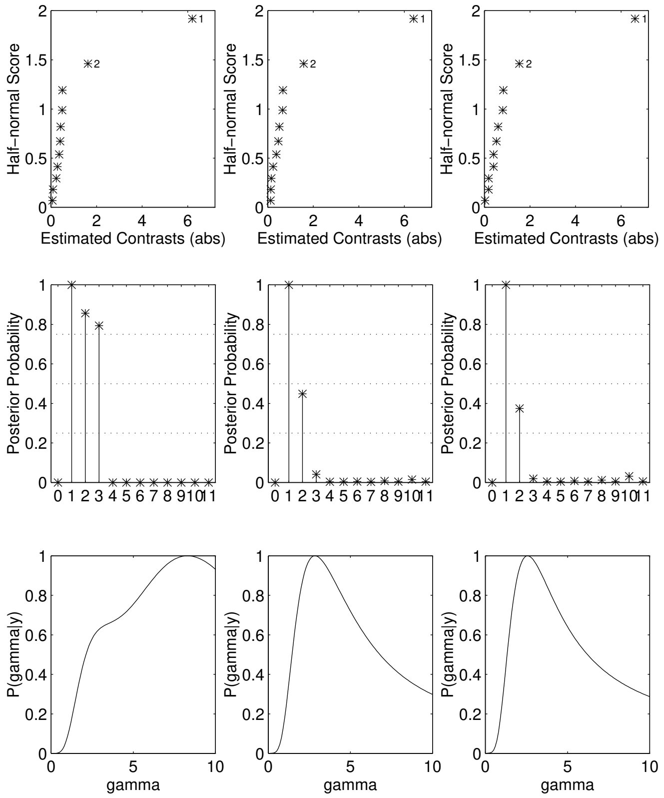

Figure 9: Half-normal plot, marginal posterior probabilities and (mode-scaled) probability density of γ. Response is given by $y = 3x_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $\epsilon \sim N(0, c^2)$ and $c = 0.2, 0.4, 0.6$ using the 12 run Plackett-Burman design.

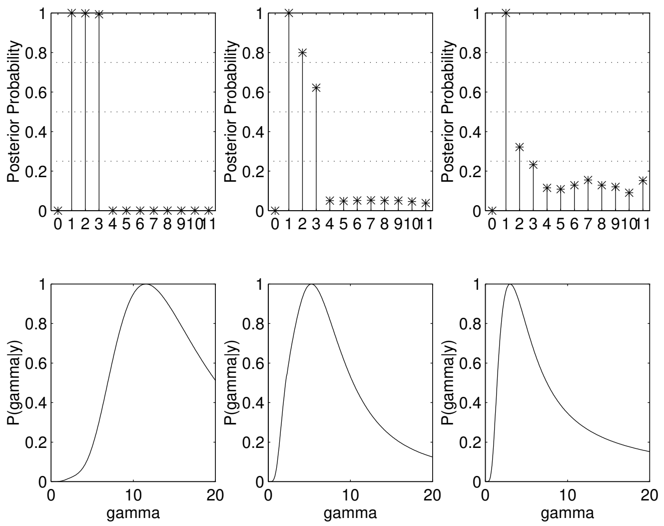

Figure 10: Half-normal plot, marginal posterior probabilities and (mode-scaled) probability density of γ for the modified Box-Meyer method. Response is given by $y = 3x_1 + x_2 + 0.5x_1x_3 + \epsilon$ with $\epsilon \sim N(0, c^2)$ and $c = 0.2, 0.4, 0.6$ using the 12 run Plackett-Burman design.

6.2 Example 3

In this section we consider an example with four active factors presented by Hamada and Wu (1992). They used a response simulated from the model $y = 2x_1 + 4x_3 + 2x_2x_3 - 2x_3x_4 + \epsilon$ with $\epsilon \sim N(0, 0.5^2)$ using the 12 run Plackett-Burman design. We will here use a realization from this model given by Hynén (1996)

$$y = (6.27\ \text{-}3.21\ 5.51\ 1.94\ \text{-}2.48\ 10.35\ 2.47\ \text{-}2.84\ \text{-}2.93\ \text{-}1.19\ \text{-}10.09\ \text{-}5.98)'.$$

Hynén analysed this example using the Box-Meyer method and concluded that, since there is a large number of models with almost equal probabilities, additional experiments are needed.

Figure 11 shows marginal posterior probabilities and (mode-scaled) probability density of gamma for this response both using the Box-Meyer method (left) and the modified Box-Meyer method (right). Models including up to four active factors are considered in the computations. Interactions of order three and higher are assumed negligible. By using the Box-Meyer method only factor x_3 stands out as being active. In contrast, the modified Box-Meyer method identifies all the four true active factors. The estimated value of γ is now 6.25 compared to 0.69 without modification.

Note that there is a drop in the probability density of γ for the modified Box-Meyer method at $\gamma = 6.248$. For $\gamma = 6.248$ the two models having highest probability are those containing the factors (x_1, x_2, x_3, x_4) and (x_1, x_2, x_4, x_5) with probabilities 0.905 and 0.074, respectively. For the last of these models only the terms $x_2, x_1x_2, x_1x_4, x_1x_5, x_2x_5$ and x_2x_5 are included. However, for $\gamma = 6.249$ also the term x_5 is included into the model. This gives a reduction in $S(\hat{\boldsymbol{\beta}})$ from 0.088 to 0.042. Due to the increased number of terms, the posterior probability of this model is however reduced from 0.074 to 0.005. In Figure 11 the value of δ is equal to 2. In Figure 12 results are shown for the modified Box-Meyer method for $\delta = 1.5$ (left) and $\delta = 2.5$ (right).

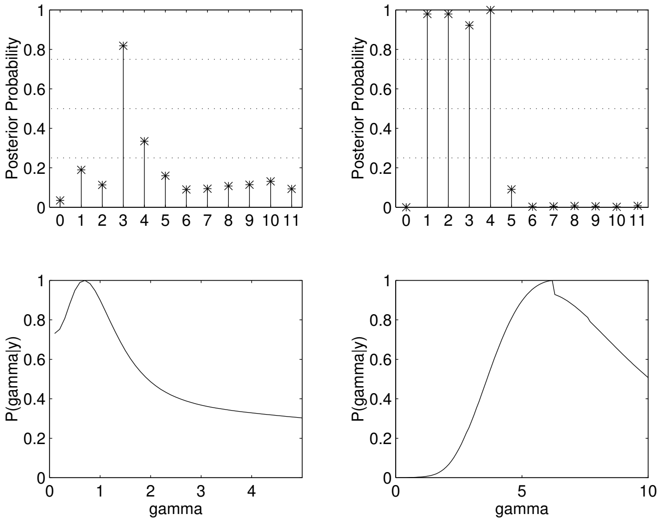

Figure 11: The marginal posterior probabilities and (mode-scaled) probability density of γ for the response $y = 2x_1 + 4x_3 + 2x_2x_3 - 2x_3x_4 + \epsilon$ with $\epsilon \sim N(0, 0.5^2)$. Left: Box-Meyer with estimated γ of 0.69. Right: Modified Box-Meyer with estimated γ of 6.25.

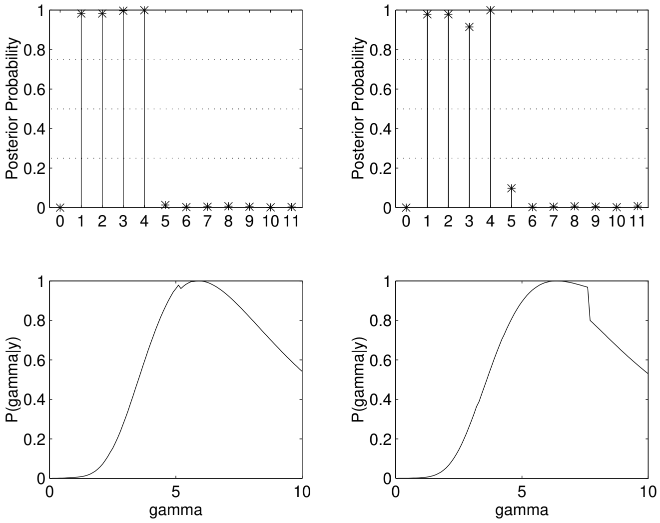

Figure 12: The marginal posterior probabilities and (mode-scaled) probability density of γ for the response $y = 2x_1 + 4x_3 + 2x_2x_3 - 2x_3x_4 + \epsilon$ with $\epsilon \sim N(0, 0.5^2)$ using the modified Box-Meyer method. Left: $\delta = 1.5$. Right: $\delta = 2.5$.

7 Summary and Conclusions

In this paper we have looked closer into the performance of the Bayesian approach for finding active factors in screening experiments proposed by Box and Meyer (1993). Using this method, we have earlier experienced somewhat more dependency of the size of both effects and error variance compared with a traditional regression analysis. These problems are illustrated using a 2_V^{5-1} design where responses are simulated from a rather simple model containing three active factors. Analysing these small examples indicate that there are too much punishment due to the number of terms in models with many potentially active factors. A simple modification of the Box-Meyer method where non-significant terms, i.e. terms with small standardised coefficients, are removed from the individual models seems to reduce the described problems. The question of how many terms to include into the models is considered by several authors in the discussion of Meyer et al. (1996), see for example the discussions by Chipman and Hamada and by Jones and DuMouchel. The results in this paper show that improvement of the Box-Meyer method can be achieved by focusing this question.

Another problem we have experienced with the Box-Meyer method is difficulty with identifying more than three active factors when using the 12 run Plackett-Burman design. An example presented by Hamada and Wu (1992) is reanalysed. The response is here simulated from a model containing four active factors using the 12 run Plackett-Burman design. An earlier analysis of these data using the Box-Meyer method has indicated that, due to a large number of almost equally probable models, additional experiments are necessary to resolve the ambiguities. We show however that the proposed modification resolves the ambiguities in this example without performing any additional experiments.

References

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. and Meyer, R. D. (1986). An Analysis for Unreplicated Fractional Factorials, *Technometrics* **28**, 11–18.

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

Chipman, H., Hamada, M. and Wu, C. F. J. (1997). A Bayesian Variable-Selection Approach for Analyzing Designed Experiments With Complex Aliasing, *Technometrics* **39**, 372–381.

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

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.

George, E. I. and McCulloch, R. E. (1993). Variable Selection Via Gibbs Sampling, *Journal of the American Statistical Association* **88**, 881–889.

Hamada, M. and Wu, C. F. J. (1991). Analysis of Censored Data From Highly Fractionated Experiments, *Technometrics* **33**, 25–38.

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). Screening for Main and Interaction Effects with Plackett and Burman's 12 Run Design, *Technical Report 1*, Division of Quality Technology and Management. Lindköping University, Sweden.

Meyer, R. D. and Box, G. E. P. (1992). Finding the Active Factors In Fractionated Screening Experiments, *Technical Report 80*, Center for Quality and Productivity Improvement. University of Wisconsin-Madison.

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.

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

Samset, O. (1998). Some Variable Selection Techniques for Finding Active Factors in Non-Geometric Plackett-Burman Designs, *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.