Discussion
Factor-based or Effect-based Modeling? Implications for Design

Hugh A. Chipman
Graduate School of Business
University of Chicago
Chicago, IL 60637

Michael S. Hamada
Bellcore
Morristown, NJ 07960

1 Introduction

We like the paper because it moves away from assuming that a single model is true when a follow-up experiment is designed. Just as there is parameter uncertainty with a limited amount of data, there can be model uncertainty. When many effects are explored using a small design, there is bound to be considerable uncertainty regarding which model best describes the data. By presenting a criterion that incorporates this uncertainty, the authors have made the theory of follow-up designs more robust to model uncertainty.

We shall consider an assumption made by the authors in their implementation regarding the form of the models entertained. Their treatment of the model space is "factor based" rather than "effect based", in the sense that many effects involving a certain factor are forced into the model if that factor is included in the model. We shall consider an effect based approach, in which individual effects have more freedom to be included or not. These two very different assumptions may be incorporated in the prior on the model space. This is shown in Section 2, along with calculations regarding the prior expectations of model size. An important issue in both factor-based and effect-based approaches is aliasing, which is explored in Section 3. Section 4 presents an algorithm for calculating posterior probability under an effect-based prior, and illustrates differences between the two approaches using the first example from the authors' paper. The form of the prior on the model space is found to have considerable influence on the choice of follow-up design.

2 Priors and Sparsity

The authors' approach involves some assumptions about relationships between main effects and higher order interactions. Specifically, the authors assume that for a given set of "active" main effects, all two- and three-factor interactions that can be formed from these main effects are included in the model. For example, if main effects for factors A, B, and C are active, the model must include the interaction terms AB, AC, BC, and ABC.

This assumption regarding "allowable" models will be called *effect forcing*. It can be embedded into a richer family of models via a prior specification of the model space proposed by Chipman (1996). In the following, an effect corresponds to a single column of the predictor matrix X; a column either corresponds to a main effect or is formed by taking the product of two or more main effect columns in which case corresponds to an interaction effect. A prior on the model space can be specified using products of probabilities that an effect is active. For example, let A, B and C denote main effects, and AB, AC, and BC their interaction effects. A model may be denoted by a vector $\boldsymbol{\delta}$, here equal to $(\delta_A, \delta_B, \delta_C, \delta_{AB}, \delta_{AC}, \delta_{BC})$. Each element is either 0 or 1, indicating an inactive or active effect respectively. The prior on $\boldsymbol{\delta}$ may be factored as follows, after making some simplifying assumptions:

$$\Pr(\boldsymbol{\delta}) = \Pr(\delta_A)\Pr(\delta_B)\Pr(\delta_C)\Pr(\delta_{AB}|\delta_A,\delta_B)\Pr(\delta_{AC}|\delta_A,\delta_C)\Pr(\delta_{BC}|\delta_B,\delta_C) \quad (1)$$

The assumptions made are that given δ's for main effects, the δ's for interactions are independent of each other, and only depend on the δ's of main effects that are used to form them. δ's for main effects are also assumed to be independent. Thus the probability that AB is active depends only on whether main effects for A and B are active.

Following the authors, we take $\Pr(\delta_i = 1) = \pi$ for i indexing any main effect. Terms like $\Pr(\delta_{AB}|\delta_A, \delta_B)$ require specification of four probabilities:

$$\Pr(\delta_{AB} = 1|\delta_A, \delta_B) = \begin{cases} p_{00} & \text{if } (\delta_A, \delta_B) = (0,0) \\ p_{01} & \text{if } (\delta_A, \delta_B) = (0,1) \\ p_{10} & \text{if } (\delta_A, \delta_B) = (1,0) \\ p_{11} & \text{if } (\delta_A, \delta_B) = (1,1) \end{cases}. \quad (2)$$

Taking

$$(p_{00}, p_{01}, p_{10}, p_{11}) = (0, 0, 0, p) \quad (3)$$

would correspond to the assumption that an interaction can only be active if both its corresponding main effects are active, and then with probability p. This is referred to as *strong heredity* by Chipman (1996). Effect forcing is a special case of strong heredity, with

$$(p_{00}, p_{01}, p_{10}, p_{11}) = (0, 0, 0, 1). \quad (4)$$

Active main effects A and B mean that the AB interaction *must* be included in the model. Other choices of the four probabilities would include *weak heredity*, which has only $p_{00} = 0$, allowing interactions to be active when one or more corresponding main effect is active. In what follows, we shall restrict attention to strong heredity and effect forcing.

The prior given above has an analog for higher order terms, such as the three-factor interaction ABC. In this case several possible specifications are possible.

δ_{ABC} could depend on $(\delta_A, \delta_B, \delta_C)$, or on $(\delta_{AB}, \delta_{AC}, \delta_{BC})$. The latter would allow higher order interactions to be active only if interactions of the next lowest order are active; the former would only have dependence on main effects. Due to the additional complexity that three-factor interactions would impose, we only deal with two-factor interactions in what follows.

Effect forcing and strong heredity priors may be compared via the expected number of terms that are active. If only main effects and two-factor interactions are to be considered, and there are a maximum of k main effects, we may write

$$\begin{aligned} \mathrm{E}(\#\text{ effects}) &= \mathrm{E}_f[\mathrm{E}(\#\text{ effects}|f\text{main effects active})] \\ &= \mathrm{E}_f(f + \binom{f}{2}p) \\ &= k\pi[1 + p(k-1)\pi/2]. \end{aligned} \tag{5}$$

The random variable f, the number of active main effects, is binomial with k trials and probability of success π. When f main effects are active, $\binom{f}{2}$ two-factor interactions may be active, each with probability p. Thus we expect that conditional on f active main effects, $p\binom{f}{2}$ interaction effects would be active. A similar calculation could be made if three-factor interactions are considered.

Choices of p between 0 and 1 yield an expected number of terms between $k\pi$ and $k\pi[1+(k-1)\pi/2]$. In the authors' example from Section 3 with $\pi = 0.25$, $k = 8$, they would expect $k\pi = 2$ factors to be active, but 3.75 effects to be active. If three-factor interactions are also forced to enter, the expected number of effects is 4.625. This is more than twice as many active effects as active factors. Choice of $p = 0.25$ gives 2.44 expected active effects. The assumption of sparsity on the factors implies less sparsity on the effects.

The main distinction between effect forcing and strong heredity is that effect forcing constrains the model space to a small sub-region of all possible models. Effect forcing given in (4) puts zero mass on many models, such as one containing only main effects and no interactions. By taking $p < 1$ in (3), such a model would receive positive prior mass.

Since the posterior ((1) in their appendix) can be factored as

$$\Pr(M_i|Y) \propto \Pr(M_i)\Pr(Y|M_i), \tag{6}$$

the relaxation of the effect forcing assumption will lead to a posterior that contains many more models. In terms of calculation, evaluation of $\Pr(Y|M_i)$ will remain unchanged, and only the form of $\Pr(M_i)$ will change. Calculation issues are addressed further in Section 4.

A wider choice of models may be attractive if it allows a more accurate description of the plausible models. In their first example, a model with effects C, E, H, EH would include all effects with large contrasts, but have a simpler form than a model also including the (apparently insignificant) effects CE, CH, CEH.

3 Aliasing

When effects are aliased in the initial design, additional uncertainty regarding the model is introduced. If two effects correspond to the same column for the given experimental runs, there will be no information in the data to distinguish between them. In the first example, the effect AE=CH=BF=DG has the second largest estimated coefficient. The prior can be used to reduce this uncertainty. Under both effect forcing and strong heredity priors, this effect is most likely to be labeled as either an AE or a CH interaction because of the large magnitude of main effects E and C. In order for this interaction effect to be included, both its parents must be included. Since B, D, F, and G all have small main effects, interactions involving these terms seem less likely. The prior constraint on inclusion of main effects reduces uncertainty regarding this effect.

An additional consequence of the effect forcing prior not mirrored by strong heredity is the inability to discriminate between aliased terms of different orders. When three-factor interactions are considered, they may play the role of a main effect to which they are aliased. This would occur in resolution IV designs. In the first example, the model with factors A, C, E would include the effect ACE, which is aliased to H. Because three-factor interactions are automatically included under effect forcing, there is no mechanism in that prior for indicating preference of a main effect over a three-factor interaction with which it is aliased. Only the (very coarse) measure of removing three-factor interactions from consideration entirely would allow discrimination between them and main effects. Strong heredity priors provide more flexibility, and tend to discriminate against higher order terms.

In many cases, identification of interactions between more than two factors may be too ambitions for small screening designs. When such terms are of interest, strong heredity priors provide more flexibility in dealing with them.

Aliasing of effects also means that some identical columns can be included in the regression matrix X. In example 1, the model with main effects A, C, E, and H and all two- and three-factor interactions involving these terms contains 14 columns plus an intercept. Only seven of these are distinct. In a frequentist regression, the duplicates would be removed. By putting an informative prior on the coefficients, all 14 columns can be included. For the original 16 runs, all five models receiving high posterior mass are the same, with possible duplication of some columns. Aliasing, in combination with effect forcing, makes it difficult to distinguish among these models, and the mass associated with the single unique set of X columns is divided among these five models. We shall show in Section 4 that strong heredity priors (which allow consideration of more models) actually resolve uncertainty in this case.

Although it may seem unusual to distinguish between five models which have identical X matrices for the original 16 runs, it is important that they all be included in the posterior. In this way, the uncertainty regarding these models is incorporated into the choice of follow-up design.

Effect forcing reduces the number of models that can be used to explain the data. In example 1, there are 8 main effects, 28 two-factor and 56 three-factor interactions. If no restrictions were placed on these effects, a total of $2^{8+28+56} = 2^{92} \approx 5 \times 10^{27}$ models are possible. Effect forcing puts zero mass on all but $2^8 = 256$ of these models. Strong heredity also reduces the number of models, but not as dramatically (in this case there are 2.8×10^8 models when only two-factor interactions are considered). The larger class of models necessitates a different computational strategy; we outline a proposal in the next section.

4 An Algorithm for Effect-based Calculations

The algorithm presented by the authors utilizes an exhaustive search that evaluates the posterior probability of all possible models in the posterior. It is made feasible by effect forcing, which greatly reduces the number of models considered. In this section, we outline a computational strategy for analogous calculations under a strong heredity prior. Key elements are utilization of aliasing structure to reduce search complexity and a non-exhaustive search. Section 4.1 outlines the algorithm; it is then applied to an example in Section 4.2.

4.1 The Algorithm

The size of the model space under strong heredity means that exhaustive calculation is intractable in all but the simplest cases. We propose a strategy that uses best subsets regression to identify promising models, and then maps these models into those in the posterior using aliasing considerations. Our emphasis is on a method that is easy to program, and runs quickly. The algorithm may be summarized as follows:

1. Use best subsets regression to identify the best m models of each size, using all allowable effects (without duplicate columns). It may be practical to ignore all models with more than a certain number of terms.

2. Delete those models from Step 1 which do not obey strong heredity.

3. For the remaining models, use the alias structure to determine the number of different models that would correspond to the same regression matrix.

4. Evaluate the posterior (up to a proportionality constant) for each of these models.

5. Renormalize the posterior on models conditional on these models.

6. Discard all models with posterior probability less than some cutoff (e.g. 0.005) - this is optional, but simplifies subsequent computation. If any are discarded, renormalize the posterior again.

7. Calculate the MD criterion over the possible designs, using the models chosen in Step 6.

In Step 1, only unique columns are considered, reducing complexity. For every model found in Step 1 that obeys strong heredity, there will be one or more corresponding model in the posterior, depending on the aliasing structure. The next section illustrates this with an example.

4.2 An Example

We consider the example used in Section 3 of the authors' paper. For simplicity, three-factor interactions are assumed to be inactive. There are 8 main effects and 28 two-factor interactions, the latter occurring in 7 alias groups each of size 4.

We first illustrate how the algorithm identifies models in the posterior that receive high mass. In this case, there are 15 unique effects to consider. The best subsets regression in Step 1 ignores all relations between these effects, and only identifies sets of effects that have high R^2. Since the shrinkage of the coefficients is symmetric, models with high R^2 and a certain number of effects will correspond to those with high mass. The model with the largest $\Pr(Y|M)$ in (6) contains the three effects $C, E, AE = CH = BF = DG$. This model is identified in Step 1, but discarded in Step 2 because it does not obey strong heredity.

Another model identified by Step 1 contains five effects: $A, C, E, H, AE = CH = BF = DG$. It obeys strong heredity, since the interaction could be between A and E, both of which are active. The alias structure suggests that there are actually two possible models with this identical regression matrix: A, C, E, H, AE and A, C, E, H, CH. Step 3 also identifies a third model corresponding to the same X matrix, but with a duplicate column: A, C, E, H, AE, CH. Models involving BF or DG terms are not possible since they do not obey strong heredity. Posteriors are evaluated for these three models, along with all others identified by Steps 1-3. The two models without duplicate columns receive equal posterior mass, since $\Pr(M_i)$ is identical for both, and $\Pr(Y|M_i)$ is identical (due to the identical X matrices). In Step 6, a cutoff of 0.02 is used;

Model	Posterior Probability
C, E, H, CH	0.4878
C, E, H, CH, EH	0.1171
A, C, E, AE	0.0942
C, E, H, CE, CH	0.0419
A, C, E, H, AE	0.0235
A, C, E, H, CH	0.0235
total	0.7878

Table 1: Posterior probability of the six most probable models with strong heredity prior for the injection molding example.

the A, C, E, H, AE, CH model is not retained because its posterior mass is too small.

In calculation of the posterior probabilities, we find that only six models have more than 0.02 posterior mass. This specific cutoff is used because all resultant models contain only terms involving A, C, E, H. A smaller cutoff would involve other terms, and make comparison with the authors' follow-up design impossible. The posterior probabilities of these six models are given in Table 1. We chose $\gamma = 2$ and $\pi = \Pr(\text{a specific main effect active}) = 0.25$ to facilitate comparison with the authors' example, and $p = 0.25$. The best subsets regression returned the 100 best models for each of sizes 1 to 10.

Paradoxically, allowing a richer class of models results in a posterior which is more concentrated. Over half the mass is on a single model: one with the three largest effects (C, E, CH), plus the necessary terms (here only H) to allow strong heredity. The marginal posteriors for individual effects are quite different from those using the authors' approach; a notable case is the absence of any terms involving the factor A in the two most probable models, meaning that the marginal probability that A is active is less than 0.4. This is in strong contrast with the value of 0.7 given for factor A by the authors in their Figure 2.

The models found under strong heredity can be contrasted with those for effect forcing. The smallest effect forcing models with considerable posterior mass have three main effects, and a total of seven terms (including three two-factor interactions, and one three-factor interaction). Effect forcing means that the smallest model containing these the three largest effects (C, E, CH) has main effects C,E,and one of A or H. Adding corresponding two- and three-factor interactions yields a seven term model. With strong heredity, the model C, E, H, CH is closer to the minimal model with only C, E, and AE=CH=BF=DG. Thus we see that by relaxing assumptions about higher order terms, a model that more exactly describes the process can be obtained.

Rank	Design Points				MD
1	11	12	15	16	8.93
2	11	11	15	15	8.71
3	11	15	15	16	8.71
4	11	11	12	15	8.66
5	9	11	15	15	8.50
⋮					
8 (7)	9	12	15	16	8.43
10 (3)	9	11	12	15	8.39
49 (4)	9	11	12	12	7.10
63 (1)	9	9	12	15	6.94
82 (2)	9	12	14	15	6.79
229 (9)	9	12	13	15	6.01
245 (10)	9	12	12	14	5.94
258 (5)	9	9	12	12	5.89
288 (8)	9	11	12	14	5.81
316 (6)	9	9	11	12	5.75
⋮					
2223(1242)	3	8	9	14	2.03

Table 2: MD criterion under strong heredity posteriors. Five best designs are followed by ten designs which were best under effect forcing, and the BHH design. The previous ranks for these 11 designs under effect forcing are given in brackets.

Table 2 gives the MD criterion evaluated with the strong heredity posterior, after conditioning on the six models given in Table 1, and renormalizing the posterior. The top five designs are given, along with the MD criterion for the 10 designs identified by the authors, and the BHH design. There are substantial differences in the ranking of designs. None of the new five best designs were in the best ten under effect forcing. Some of the effect forcing designs are still near optimal, but others have dropped to rank 200 or more below the optimal designs. Similarities between the two rankings are also evident. Rows 9, 11, 12, and 15 still dominate, and none of the designs favor re-use of the original 8 runs. The BHH design is very sub-optimal under both posteriors.

This calculation illustrates that although the MD criterion is more robust to model uncertainty, like many Bayesian methods, it is sensitive to prior assumptions. Effect forcing and strong heredity priors represent quite different assumptions about models, and this difference manifests itself in both the posterior distributions and the resultant MD criterion.

References

Chipman, H. (1996), "Bayesian Variable Selection with Related Predictors", *Canadian Journal of Statistics*, 24, 17–36.