Bayesian Analysis of Location and Dispersion Effects in Unreplicated Experiments

Russell V. Lenth
Department of Statistics
University of Iowa
Iowa City, IA 52242

July 28, 1999

Abstract

Box and Meyer (1986a) presents a Bayesian method for analyzing location effects in orthogonal unreplicated designs. Their method is based on an effect-sparsity prior that asserts that only a low proportion of effects are likely to be active. However, the variance of the response is assumed constant across all factor combinations. In many industrial experiments, this is not a safe assumption, and in fact an understanding of dispersion may be of interest in its own right. In this paper, we discuss the simultaneous study of location and dispersion effects, where there is an effect-sparsity prior for both. Examples emphasize small, unreplicated experiments; but in fact the methods place no inherent restrictions on whether the experiment is balanced or replicated.

1 Introduction

In this article, we consider a Bayesian model for simultaneous location and dispersion effects in experimental data. Such a model is especially important in industrial experiments, which are often quite small due to cost and time considerations, and where understanding of the factors that cause variation are as important or even more important than what can cause a shift in the mean.

The starting point for this work is the model proposed in Box and Meyer (1986a), which places an *effect sparsity* prior on the location effects in a balanced, unreplicated experiment. By this we mean that we have a prior conception that most location effects are zero or "inactive," and only a few are nonzero or "active." The effect-sparsity prior specifies that the active effects are independent draws from a normal distribution. The probability that an effect is active is specified and held fixed, and the prior variance of the active effects is a specified multiple of the error variance.

In a companion paper, Box and Meyer (1986b) discuss dispersion effects in the same types of experiments; however, their method is rather *ad hoc* and does not rely on an analogue of the Bayesian effect-sparsity model that they used for location. Such an approach was not tractable at the time. Now we have a rapidly expanding literature on Markov Chain Monte Carlo (MCMC) methods, which make it possible to study much more complicated Bayesian models than is possible analytically. The gist of it is that we simulate parameter values from the posterior distribution (a good general reference on MCMC methods is Gilks et al. (1996)). The purpose of this article is to adapt the effect-sparsity model in Box and Meyer (1986a) to dispersion ef-

fects, using MCMC techniques.

There is a vast literature on the Bayesian location model, which is inexorably linked with Bayesian variable selection. Among the important references are George and McCulloch (1993); Carlin and Chib (1995); George and McCulloch (1997); Chipman (1998), which use Gibbs sampling to extend Box and Meyer's work in various ways, including unbalanced data. Chipman (1996); Chipman et al. (1997) provide important extensions by imposing more restrictive effect-sparsity priors that impose various types of hierarchy among the predictors; for example, a two-way interaction can have prior probability of zero, $\pi 1$, or π_2 depending on how many of the main effects it contains are present in the model.

Important references for dispersion estimation include Bergman and Hynén (1997), which proposes a competitor to the Box and Meyer (1986b) approach. Nelder and Lee (1991); Engel and Huele (1996); Nelder and Lee (1998) discuss (and debate) generalized linear models for simultaneous location and dispersion. Mays and Myers (1996) considers a two-stage Bayesian model. Wolfinger and Tobias (1998) treat a very general setting with location and dispersion effects in complex designs, using mixed-model methods.

We define the basic location-dispersion model in Section 2. Computational issues are described in Section 3, particularly the reversible-jump method (Green, 1995) that is necessary to handle the varying dimensionality of the parameter space. Two illustrative examples are given in Section 4—a standard example used in much of the literature, and a supersaturated design. Section 5 outlines some possible extensions, and some technical details are relegated to appendices.

2 Model and priors

Assume that we have n response values $y_1, y_2, \ldots, y_n$, independent and normally distributed with

$$\mathrm{E}(y_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \tag{1}$$

$$\log(\mathrm{Var}(y_i)) = \log \sigma_i^2 = \log \sigma_\epsilon^2 + \sum_{j=1}^{q} \gamma_j z_{ij} \tag{2}$$

where the x_{ij} are predictors for location effects, and the z_{ij} are predictors for dispersion effects. (Some or all of the x_{ij} may be identical to the z_{ij}, but this is not required.) For identifiability, we impose the constraints:

$$\sum_{i=1}^{n} z_{ij} = 0, \quad j = 1, 2, \ldots, q \tag{3}$$

Thus, σ_ϵ^2 is the geometric mean of all the σ_i^2, independently of the γ_j. It is also desirable, but not necessary, to center the x_{ij}s and to scale both the x_{ij}s and the z_{ij}s so that they all have the same variance. Such additional constraints will improve the computation and make the effect-sparsity priors (discussed below) more meaningful.

It is convenient to express this model in vector form:

$$\boldsymbol{y} \sim N(\beta_0 \mathbf{1}_n + \boldsymbol{X}\boldsymbol{\beta},\ \sigma_\epsilon^2 \boldsymbol{W}^{-1}) \tag{4}$$

where $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]'$, $\boldsymbol{X}$ is the $n \times p$ matrix with elements x_{ij}, $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_p]'$, $\log \boldsymbol{W}^{-1}$ is the diagonal matrix with elements $\boldsymbol{Z}\boldsymbol{\gamma}$, $\boldsymbol{Z}$ is the $n \times q$ matrix with elements z_{ij}, and $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_q]'$. Subsequently, $\boldsymbol{x}_j$ denotes the jth column of $\boldsymbol{X}$, and $\boldsymbol{z}_j$ denotes the jth column of $\boldsymbol{Z}$.

Similar to Box and Meyer (1986a), we place effect-sparsity priors on the location and dispersion effects:

$$\beta_0 \sim N(0, d^2\sigma_\epsilon^2) \tag{5}$$

$$\beta_j \sim \alpha N(0, c^2\sigma_\epsilon^2) + (1-\alpha)I_0, \quad j = 1, 2, \ldots, p \tag{6}$$

$$\gamma_j \sim \phi N(0, \sigma_\gamma^2) + (1-\phi)I_0, \quad j = 1, 2, \ldots, q \tag{7}$$

all independently, where I_0 denotes the degenerate distribution at 0. Note that α and ϕ specify the respective probabilities that a location or dispersion effect are "active" or nonzero. c^2 is the ratio of the prior variance of an active location effect to the error variance; this is equivalent to the formulation in Box and Meyer (1986a), except that they work in terms of a parameter $k^2 = 1 + nc^2$. The variance σ_γ^2 does not depend on the scale of the data; instead, it controls the relative variations in the σ_i^2. Finally, note that β_0 is taken to be always active.

Additional priors are placed on the variances:

$$1/\sigma_\epsilon^2 \sim \Gamma(\nu, \psi) \tag{8}$$

$$\sigma_\gamma \sim U(0, \lambda) \tag{9}$$

where $\Gamma(\nu, \psi)$ is the gamma distribution with mean $\nu\psi$ and variance $\nu\psi^2$. This is the conjugate prior for σ_ϵ. The rationale for choosing a bounded prior on σ_γ is discussed in Section 3.

The remaining parameters α, ϕ, c, and d are "hyper parameters," i.e., parameters of the priors already described. We may hold these fixed at prescribed values, or put prior distributions on them.

In the Box and Meyer (1986a) formulation for the location model, the intercept β_0 is removed from the problem, which is equivalent to setting

$d = \infty$. In addition, they use a vague prior on σ_ϵ with density proportional to $1/\sigma_\epsilon$; this is equivalent to using $\nu = 0$ and $\psi = \infty$. They hold α and c fixed, recommending $\alpha = .20$ and $k = \sqrt{1 + nc^2} = 10$. Haaland and O'Connell (1995) point out that k^2 is like a population F ratio for the active effects, and recommend $\alpha = .4$ and $k = 5$.

3 Computation

We now discuss the particulars of the MCMC sampler. The joint posterior density is derived in Appendix A; it depends on the indicators $\boldsymbol{\delta} = [\delta_1, \delta_2, \ldots, \delta_p]'$, $\boldsymbol{\omega} = [\omega_1, \omega_2, \ldots, \omega_q]'$ where

$$\delta_j = 1_{\beta_j \neq 0}, \quad j = 1, 2, \ldots, p \tag{10}$$

$$\omega_j = 1_{\gamma_j \neq 0}, \quad j = 1, 2, \ldots, q \tag{11}$$

Our primary focus is on studying which effects are active, as measured by the posterior probabilities that $\delta_j = 1$ or $\gamma_j = 1$. It is possible to simplify the problem somewhat by integrating out $\boldsymbol{\beta}$ and σ_ϵ, yielding a marginal posterior pdf that depends principally on the residual sum of squares

$$S(\boldsymbol{\delta}, \boldsymbol{\gamma}) = (\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\hat{\tilde{\boldsymbol{\beta}}})'\tilde{\boldsymbol{W}}\tilde{\boldsymbol{y}} \tag{12}$$

for an augmented linear system of the form $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\epsilon}}$, with least-squares estimator $\hat{\tilde{\boldsymbol{\beta}}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{W}}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{y}}$. Details are given in Appendix A; the linear model is specified in (20), and the marginal posterior is given in (23).

3.1 Location-model parameters

After integrating out $\boldsymbol{\beta}$ and σ_ϵ, the remaining parameters in the location model are $\boldsymbol{\delta}$, and possibly some or all of c, d, and α, if these are not held fixed. We may use ordinary Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990) to simulate the δ_j. In Gibbs sampling, parameter values are simulated from the full conditional distribution, i.e., the distribution of the parameter in question, given the current values of all other parameters. By inspection, it is easy to verify that the distribution of δ_j, given all other parameters, is Bernoulli with parameter $P_j = R_j/(1+R_j)$, where

$$R_j = \frac{1}{c} \cdot \frac{\alpha}{(1-\alpha)} \left(\frac{\det(\tilde{\boldsymbol{X}}'_{-j}\tilde{\boldsymbol{X}}_{-j})}{\det(\tilde{\boldsymbol{X}}'_{+j}\tilde{\boldsymbol{X}}_{+j})} \right)^{1/2} \left(\frac{2/\psi + S(\boldsymbol{\delta}_{-j}, \boldsymbol{\gamma})}{2/\psi + S(\boldsymbol{\delta}_{+j}, \boldsymbol{\gamma})} \right)^{\nu+n/2} \tag{13}$$

and the subscripts $-j$ and $+j$ indicate that δ_j is set to 0 or 1, respectively (i.e., $\boldsymbol{x}_j$ is removed or included in the location model). The necessary quantities can be computed efficiently using a rank-one update of the linear system (Gill et al., 1974, 1975).

In the simplest case, the remaining location-model parameters c, d, and α are held fixed at specified values. If instead, α is assigned a beta prior with parameters (a_α, b_α), then α may be generated using Gibbs sampling; its full conditional distribution is beta with parameters $(a_\alpha + m_\beta, b_\alpha + p - m_\beta)$ where $m_\beta = \mathbf{1}'\boldsymbol{\delta}$, the number of active β_js.

There is no obvious easy choice for a prior on c or d. If a prior pdf $\xi_c(c)$ is placed on c, the Metropolis-Hastings (Hastings, 1970) algorithm may be applied. We generate a candidate value c^* from a specified "proposal" density

$\tau(c^*|c)$, and replace $c \leftarrow c^*$ with probability $\min(1, P_c)$, where

$$P_c = \frac{\xi_c(c^*)\tau(c|c^*)}{\xi_c(c)\tau(c^*|c)} \left(\frac{c}{c^*}\right)^{m_\beta} \left(\frac{\det(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})}{\det(\tilde{\boldsymbol{X}}^{*\prime}\tilde{\boldsymbol{X}}^*)}\right)^{1/2} \left(\frac{2/\psi + S(\boldsymbol{\delta},\boldsymbol{\gamma})}{2/\psi + S^*(\boldsymbol{\delta},\boldsymbol{\gamma})}\right)^{\nu+n/2} \quad (14)$$

Here, the $*$ superscript indicates that the quantity in question is computed using c^* in place of c. The method for d is of identical form to (14), with c replaced by d and 1 in place of m_β. In the particular case where c has a $\Gamma(a, b)$ prior and $c^*|c \sim \Gamma(f, c/f)$, we have

$$\begin{aligned} P_c &= \left(\frac{c}{c^*}\right)^{m_\beta+2f-a} \exp\left\{\frac{c-c^*}{b} + \left(\frac{c^*}{c} - \frac{c}{c^*}\right) f\right\} \\ &\quad \times \left(\frac{\det(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})}{\det(\tilde{\boldsymbol{X}}^{*\prime}\tilde{\boldsymbol{X}}^*)}\right)^{1/2} \left(\frac{2/\psi + S(\boldsymbol{\delta},\boldsymbol{\gamma})}{2/\psi + S^*(\boldsymbol{\delta},\boldsymbol{\gamma})}\right)^{\nu+n/2} \end{aligned} \quad (15)$$

3.2 Dispersion-model parameters

The sampler for the $\boldsymbol{\gamma}$ (and its associated indicators, $\boldsymbol{\omega}$) is somewhat complicated because the dimension of the parameter space changes when a nonzero γ_j is set to zero, or vice versa. The solution is to use a reversible-jump algorithm (Green, 1995). A suitable scheme is specified by the following algorithm; details of the derivation are given in Appendix B.

Algorithm ARC

Define moves A, R, and C as follows:

A (Add $\boldsymbol{z}_j$ to the model)

Generate a new $\gamma_j \sim N(0, \sigma_\gamma^2)$. With probability $\min(1, A_{+j})$ (see (16) below), accept this new value; otherwise, let γ_j revert to its previous value.

R (Remove $\boldsymbol{z}_j$ from the model)

With probability $\min(1, 1/A_{+j})$, set $\gamma_j \leftarrow 0$.

C (Change the value of γ_j)

Let $\gamma_j^* = \gamma_j + D_j$ where $D_j \sim N(0, \sigma_D^2)$. With probability $\min(1, B_j)$, (see (17) below) set $\gamma_j \leftarrow \gamma_j^*$; otherwise, leave γ_j unchanged.

Select from among these moves as follows:

- If $\gamma_j = 0$, choose Move A.
- If $\gamma_j \neq 0$, choose Move R with probability P_R; otherwise, choose Move C.

In this algorithm, P_R and σ_D are specified to provide good mixing properties (we suggest $P_R = .5$ and $\sigma_D = 0.1$), and

$$A_{+j} = P_R \cdot \frac{\phi}{1-\phi} \left(\frac{2/\psi + S(\boldsymbol{\delta}, \boldsymbol{\gamma}_{-j})}{2/\psi + S(\boldsymbol{\delta}, \boldsymbol{\gamma}_{+j})} \right)^{\nu+n/2} \tag{16}$$

$$B_j = \frac{\exp\{-\frac{1}{2}(\gamma_j^*/\sigma_\gamma)^2\}}{\exp\{-\frac{1}{2}(\gamma_j/\sigma_\gamma)^2\}} \left(\frac{2/\psi + S(\boldsymbol{\delta}, \boldsymbol{\gamma})}{2/\psi + S(\boldsymbol{\delta}, \boldsymbol{\gamma}^*)} \right)^{\nu+n/2} \tag{17}$$

where the subscript $-j$ and $+j$ denote that computations are carried out with γ_j set respectively to zero or to the nonzero value specified in the move; and $\boldsymbol{\gamma}^*$ is the same as $\boldsymbol{\gamma}$ except its jth element is set to γ_j^*. Note that, in comparing candidate values of γ_j, we are varying the weights assigned to observations. This affects the value of $\hat{\tilde{\boldsymbol{\beta}}}$ and so a linear system must be re-solved. Unfortunately, this requires more computation than does the rank-one update we use in the location model.

The value of σ_γ is updated using a Metropolis-Hastings step. Generate $G \sim U(0, \lambda)$ and let

$$R = \left(\frac{\sigma_\gamma}{G}\right)^{m_\gamma} \frac{\exp(-\frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{\gamma}/G^2)}{\exp(-\frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{\gamma}/\sigma_\gamma^2)} \tag{18}$$

where $m_\gamma = \mathbf{1}'\boldsymbol{\omega}$ is the current number of nonzero γ_js. With probability $\min(1, R)$, we set $\sigma_\gamma \leftarrow G$, otherwise we leave it alone. Note that when $m_\gamma = 0$, $R = 1$, forcing a new σ_γ to be generated.

If we wish to put a prior on ϕ rather than holding it fixed, the easiest case is a beta prior with parameters (a_ϕ, b_ϕ); then the full conditional is beta with parameters $(a_\phi + m\gamma, b_\phi + q - m_\gamma)$ and Gibbs sampling may be used.

3.3 Summary of the MCMC method

One iteration of the MCMC sampler involves giving every parameter a chance to change:

1. For $j = 1, 2, \ldots, p$: Generate δ_j using Gibbs sampling and (13).

2. For $j = 1, 2, \ldots, q$: Generate γ_j (and hence ω_j) using the reversible-jump scheme in Algorithm ARC.

3. Generate σ_γ using the Metropolis-Hastings method, (18) and associated discussion.

4. If priors are used for α, c, d, or ϕ, generate new ones using suitable Gibbs or Metropolis-Hastings steps, as specified above.

All computations for updating one parameter are performed based on the current values of all other parameters.

Several thousand iterations should be used in hopes of gaining thorough coverage of the posterior distribution. The first few hundred iterations are usually ignored because the starting parameter values may not be near a dense region. As a basic starting point, we use $\boldsymbol{\gamma} = \mathbf{0}$ and set $\delta_j = 1$ only for those j where $|(\boldsymbol{y} - \overline{y}\mathbf{1})'\boldsymbol{x}_j|$ exceeds the average of these quantities. It is worth trying several runs with different starting points, to see if the results are comparable. There is a growing literature on convergence diagnostics for MCMC methods; see, for example, Cowles and Carlin (1996); Brooks and Roberts (1999); Mengersen et al. (1998). Convergence assessment is especially complicated in model-selection problems, such as this one, where the dimension of the parameter space changes. Little is published in this area, but Brooks and Giudici (1999) provides some guidance.

With good algorithms and compiled code, the sampler described is not unduly burdensome. For example, using a compiled C program on the author's Hewlett-Packard 9000/777 unix workstation, it takes only a few seconds to simulate 10,000 iterations when $n = 50$, $p = 15$, and $q = 15$.

3.4 Why a bounded prior is needed

We use a uniform prior on σ_γ instead of an inverted gamma (which is conjugate) so as to avoid some numerical difficulties. Early test runs using the inverted gamma prior in the MCMC sampler often produced a situation where σ_γ diverged and $S(\boldsymbol{\delta}, \boldsymbol{\gamma})$ approached zero. When σ_γ is very large, the $|\gamma_j|$ grow, producing near-zero and near-infinite weights in $\boldsymbol{W}$. This effectively discards some of the data. Intuitively, it appears that with enough random exploration and enough location and dispersion predictors, one can

often settle on a location/dispersion model where only a handful of residuals are nonzero, and where those observations are assigned extremely low weights.

To avert this phenomenon, it is apparent that we need to put limits on the dispersion model so that every observation is guaranteed at least a small amount of weight. One way to do this is to simply truncate the support of the priors for $\boldsymbol{\gamma}$. It seems more elegant, however, to bound the distribution of σ_γ, thereby producing smoother posterior densities of the γ_j without hard truncation.

As the number q of dispersion effects grows, the value of λ must decrease in order to be effective in limiting the weights. On the other hand, making λ too small could prevent important dispersion effects from being noticed. We have found that $\lambda = 5/\sqrt{q}$ usually works well.

4 Examples

4.1 Welding-strength data

The main example in Box and Meyer (1986b) (also used as Example II in Box and Meyer (1986a)) concerns a welding-strength experiment in 16 runs and 13 factors (9 main effects and 6 two-way interactions). The data, listed in Table 1, were reported in Taguchi and Wu (1986) and are studied by several other authors, including Bergman and Hynén (1997). Figure 1 shows the results of two MCMC simulations of 5,500 iterations with $d = \infty$, $\nu = \psi = 0$, $\phi = .2$, and $\lambda = 5/\sqrt{13}$, using all 13 predictors in both the location and the dispersion model. Here and elsewhere, the first 500 iterations are

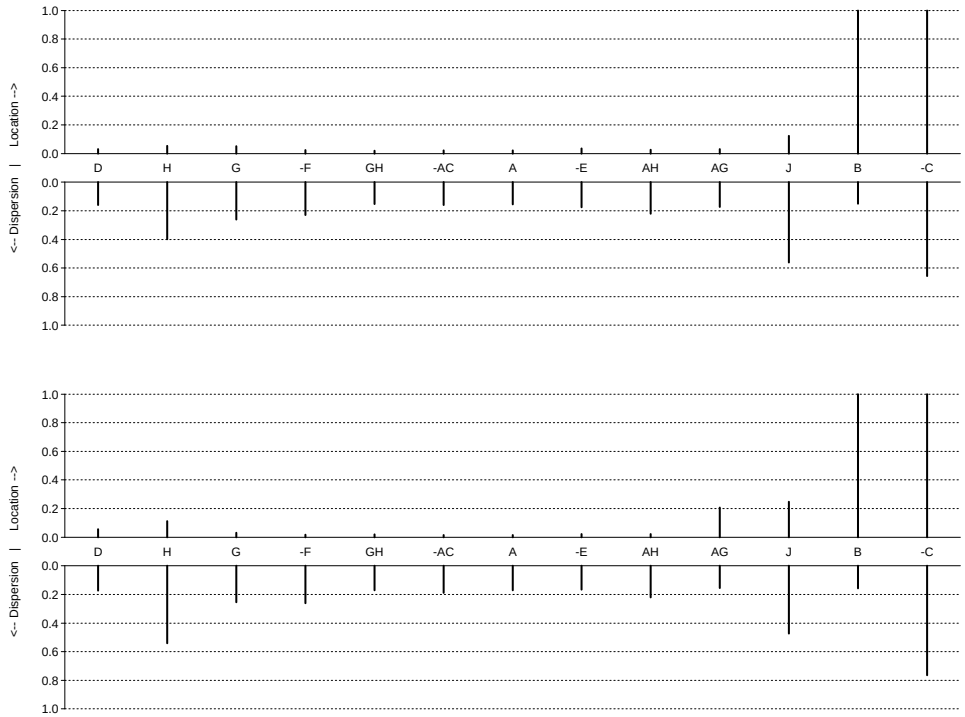

Figure 1: Posterior probability plots for the welding data. The top set shows location and dispersion effects using $c = 2.5$, $\alpha = \phi = .2$, $\lambda = 5/\sqrt{15}$. The bottom set uses a Beta(3,12) prior on α and a Gamma(10,.25) prior on c, with the same fixed ϕ and λ.

excluded from the summaries. The first run uses fixed $\alpha = .2$ and $c = 2.5$ ($k = \sqrt{1 + 16c^2} \approx 10$), the values recommended in Box and Meyer (1986a). In the second run, we allowed α and c to vary. The prior for α is Beta with parameters $(3, 12)$, and the prior for c is Gamma with shape parameter 10 and scale parameter 0.25. From the last 5,000 simulations, the posterior distribution of α had quartiles $(.15, .21, .26)$, and c had quartiles $(2.6, 3.3, 3.9)$.

The plots represent the simulated posterior probabilities that $\delta_j = 1$ (upper bars) and $\omega_j = 1$ (lower bars). The results are comparable to those obtained by other authors: by far the most prominent location effects are the main effects of factors B and C, and the most prominent dispersion effects

Table 1: The welding-strength data. In the analysis, the codings $-$ and $+$ were replaced by -1 and $+1$.

Run	D	H	G	$-F$	GH	$-AC$	A	$-E$	AH	AG	J	B	$-C$	y
1	$-$	$-$	$-$	$+$	$+$	$-$	$-$	$+$	$+$	$+$	$-$	$-$	$+$	43.7
2	$+$	$-$	$-$	$-$	$+$	$+$	$-$	$-$	$+$	$+$	$+$	$-$	$-$	40.2
3	$-$	$+$	$-$	$+$	$-$	$+$	$-$	$+$	$-$	$+$	$-$	$+$	$-$	42.4
4	$+$	$+$	$-$	$-$	$-$	$-$	$-$	$-$	$-$	$+$	$+$	$+$	$+$	44.7
5	$-$	$-$	$+$	$-$	$-$	$+$	$-$	$+$	$+$	$-$	$+$	$+$	$-$	42.4
6	$+$	$-$	$+$	$+$	$-$	$-$	$-$	$-$	$+$	$-$	$-$	$+$	$+$	45.9
7	$-$	$+$	$+$	$-$	$+$	$-$	$-$	$+$	$-$	$-$	$+$	$-$	$+$	42.2
8	$+$	$+$	$+$	$+$	$+$	$+$	$-$	$-$	$-$	$-$	$-$	$-$	$-$	40.6
9	$-$	$-$	$-$	$+$	$+$	$-$	$+$	$-$	$-$	$-$	$+$	$+$	$-$	42.4
10	$+$	$-$	$-$	$-$	$+$	$+$	$+$	$+$	$-$	$-$	$-$	$+$	$+$	45.5
11	$-$	$+$	$-$	$+$	$-$	$+$	$+$	$-$	$+$	$-$	$+$	$-$	$+$	43.6
12	$+$	$+$	$-$	$-$	$-$	$-$	$+$	$+$	$+$	$-$	$-$	$-$	$-$	40.6
13	$-$	$-$	$+$	$-$	$-$	$+$	$+$	$-$	$-$	$+$	$-$	$-$	$+$	44.0
14	$+$	$-$	$+$	$+$	$-$	$-$	$+$	$+$	$-$	$+$	$+$	$-$	$-$	40.2
15	$-$	$+$	$+$	$-$	$+$	$-$	$+$	$-$	$+$	$+$	$-$	$+$	$-$	42.5
16	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	46.5

are due to C, J (labeled "I" by some authors), and H. The results are similar for the two Bayesian models illustrated. The second model tends to show stronger posterior probabilities for a few of the location effects.

Other MCMC simulations (not shown) were performed where both α and ϕ were assigned Beta(3,12) priors. This caused some instability in the computation as is described in Section 3.4, where most of the predictors were made active in the dispersion model and $S(\boldsymbol{\delta}, \boldsymbol{\gamma})$ approached zero. Reducing λ to 0.4 stabilized things, and we then obtained results similar to those in Figure 1. We also tried various runs where a few randomly selected observations were removed. With one observation discarded, there was little change. With four discarded, the posterior probabilities for both location and dispersion effects were noticeably altered (mostly downward) but still yielded similar conclusions. This is the behavior one would expect, given a significantly smaller sample size.

4.2 A supersaturated design

Lin (1993) reports an example of a supersaturated experiment in 24 factors but only 14 runs. This is actually a half fraction of a 28-run experiment given in Williams (1968) and Box and Draper (1987), page 175. All 28 observations are listed in Table 2, with the half fraction specially marked. Using only the 14 runs of the supersaturated design, Lin's analysis (which does not include a dispersion model) identifies factors 15, 12, 20, 4, and 10 as the most important location effects. Using our methods with no dispersion model and a location model with $\alpha = .2$ and $c = 2.5$, we obtain the same top five location effects in the same order, as shown in the top panel in Figure 2.

Table 2: Data from the rubber experiment (Williams, 1968). The supersaturated design comprises the runs numbered with the superscript s.

Run	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}	y
1^s	+	+	+	−	−	−	+	+	+	+	+	−	+	−	−	+	+	−	−	+	−	−	−	+	133
2	−	+	−	−	−	−	+	+	+	+	−	+	−	−	+	−	+	+	+	+	−	+	+	−	49
3^s	+	−	−	−	−	−	+	+	+	−	−	−	+	+	+	+	−	+	−	−	+	+	−	−	62
4^s	+	+	−	+	+	−	−	−	−	+	−	+	+	+	+	+	+	−	−	−	−	+	+	−	45
5	+	+	−	−	+	+	−	−	−	−	−	−	−	−	−	−	+	+	−	+	+	+	−	+	88
6^s	+	+	−	+	−	+	−	−	−	+	+	−	+	−	+	+	−	+	+	+	−	−	−	−	52
7	−	−	+	+	+	+	+	+	−	+	−	+	+	−	−	+	−	+	−	−	−	−	−	−	300
8^s	−	−	+	+	+	+	−	+	+	−	−	−	+	−	+	+	+	−	−	+	−	+	+	+	56
9^s	−	−	+	+	+	+	+	−	+	+	+	−	−	+	+	−	+	+	+	+	+	+	−	−	47
10^s	−	−	−	−	+	−	−	+	−	+	−	+	+	+	−	+	+	+	+	+	+	−	−	+	88
11	+	−	+	−	−	+	−	−	+	−	−	−	+	+	−	+	+	+	+	−	−	−	+	−	116
12	−	+	+	+	−	−	+	−	−	+	+	−	+	+	−	+	+	+	−	−	+	+	+	+	83
13^s	−	+	+	−	−	+	−	+	−	+	−	−	−	−	−	−	−	−	+	−	+	+	+	−	193
14	−	−	−	+	−	−	−	−	+	+	−	−	−	−	+	−	−	−	−	−	+	−	−	+	230
15	+	−	+	−	+	−	+	−	−	+	−	−	−	+	+	−	−	−	+	+	−	−	+	+	51
16	−	+	+	−	+	−	−	−	+	−	+	+	+	−	+	+	+	−	+	−	+	−	−	−	82
17^s	−	−	−	−	−	+	+	−	−	−	+	+	−	−	+	−	+	+	−	−	−	−	+	+	32
18	+	−	+	+	−	−	−	+	−	−	+	+	+	−	+	+	−	+	+	+	+	+	+	+	58
19	+	−	−	+	+	−	+	+	−	−	+	−	−	−	−	−	+	−	+	−	−	+	−	−	201
20	+	+	+	−	+	+	−	+	+	+	+	+	−	+	+	−	−	+	−	−	−	+	−	+	56
21	−	+	−	+	−	+	+	−	+	−	−	+	+	+	−	+	−	−	+	+	−	+	−	+	97
22^s	+	+	+	+	−	+	+	+	−	−	−	+	−	+	+	−	+	−	+	−	+	−	−	+	53
23^s	−	+	−	+	+	−	−	+	+	−	+	−	−	+	−	−	−	+	+	−	−	−	+	+	276
24^s	+	−	−	−	+	+	+	−	+	+	+	+	+	−	−	+	−	−	+	−	+	+	+	+	145
25^s	+	+	+	+	+	−	+	−	+	−	−	+	−	−	−	−	−	+	−	+	+	−	+	−	130
26	−	+	−	−	+	+	+	+	−	−	+	−	+	+	+	+	−	−	−	+	+	−	+	−	55
27	+	−	−	+	−	+	−	+	+	+	+	+	−	+	−	−	+	−	−	+	+	−	+	−	160
28^s	−	−	+	−	−	−	−	−	−	−	+	+	−	+	−	−	−	−	−	+	−	+	−	−	127

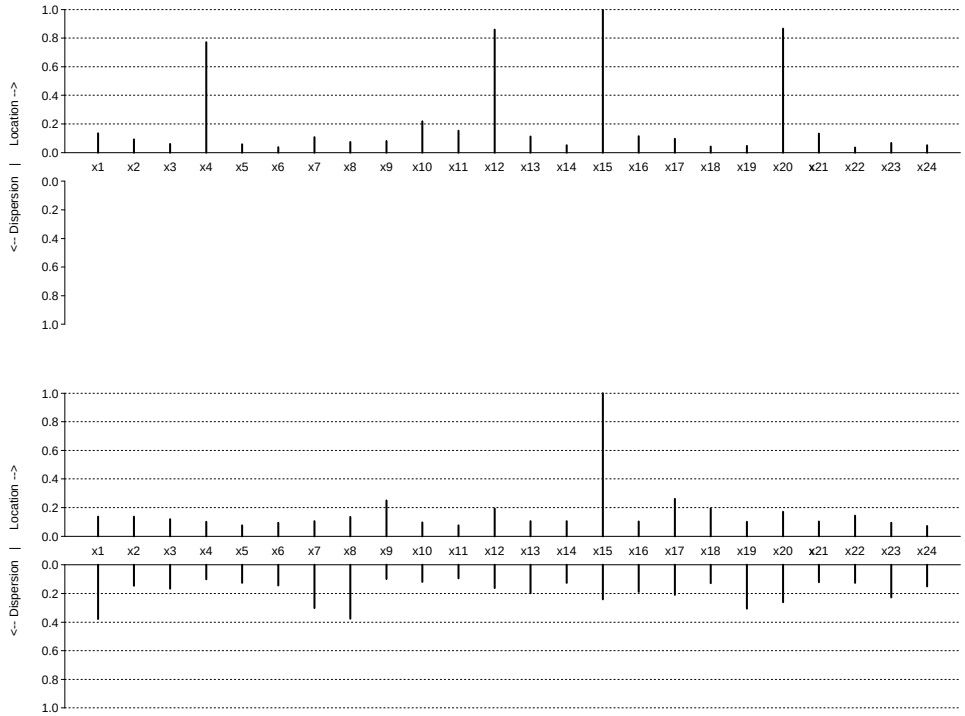

Figure 2: Posterior probability plots for the supersaturated design. The top set shows results for a location-only model with $\alpha = .2$, $c = 2.5$. The bottom set uses the same location model plus a dispersion model with $\phi = .1$, $\lambda = 5/\sqrt{24}$.

Factor 10 is not nearly as prominent as the other four; this is also consistent with Lin's analysis. As before, we have summarized the last 5,000 of 5,500 iterations of the MCMC sampler.

Now we add a dispersion model. We have already seen in the welding example that dispersion effects tend to come out with somewhat high posterior probabilities; so we have elected to suppress them more than the location effects by setting $\phi = .1$. As before, we use $\lambda = 5/\sqrt{24} \approx 1$. The result of including all 24 location predictors and all 24 dispersion predictors is depicted in the bottom panel of Figure 2. What is interesting here is the effect on the location model: three of the four big location effects have completely lost their prominence, achieving even lower posterior frequencies than some other factors.

While the presence of dispersion effects has had a big impact, there are no clear winners among them. Figure 3 shows the results from two simplified models. The first one includes all of the predictors in the location model, and the top five dispersion predictors. From this model, we selected the top three location and dispersion predictors, yielding the results in the bottom panel. The parameter λ in each of these models is set at its software default values of $5/\sqrt{5} \approx 2.2$ and $5/\sqrt{3} \approx 2.9$, respectively. It appears that much of the data can be explained very simply in terms of one location effect (factor 15) and one or two dispersion effects (factor 1 and perhaps factor 8). It is important to note that these are probably not the only reasonable, parsimonious models that can be fitted to these data; but they are appealing, and can provide some guidance for further experimentation.

One advantage of the MCMC approach is that the simulation results

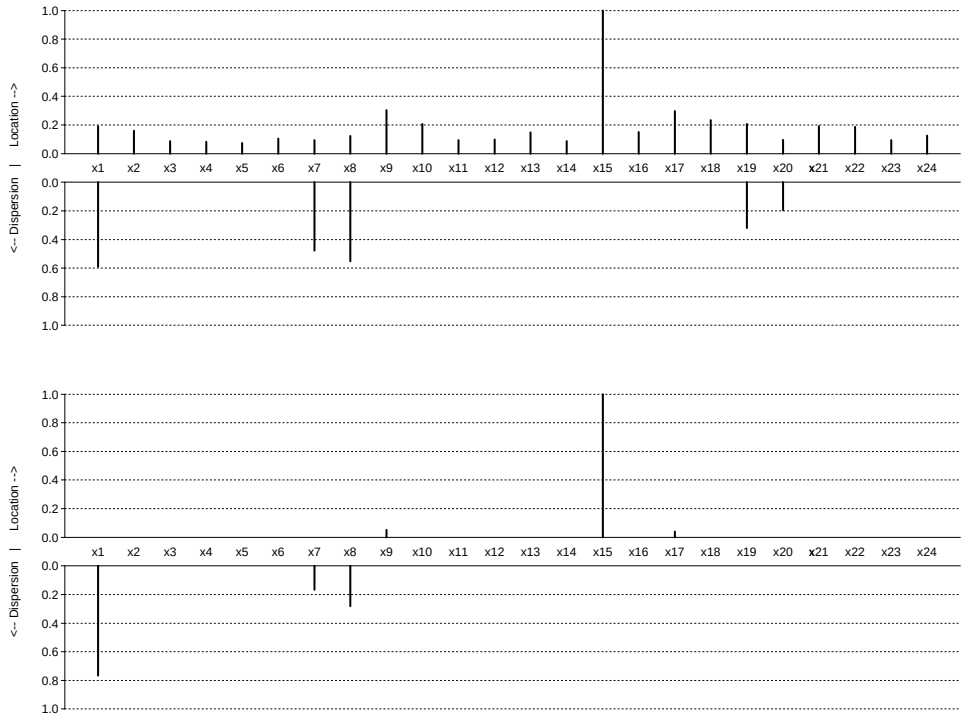

Figure 3: Posterior probability plots for reduced models for the supersaturated example. The top set includes all location effects and the 5 most prominent dispersion effects from the bottom panel of Figure 2. The bottom set further reduces to the three most prominent location and dispersion effects.

themselves can be useful to gain more detailed insight. In effect, the sampler behaves like a hyperactive research assistant who, with minimal guidance provided by the prior specifications, fits thousands of different models to the data to see what works the best. By exploring the simulation output, one can discern what factors tend to appear together in the same model and render a more complete picture than what is summarized here.

We can also look at the simulated parameter values. This must be done with care, however, to avoid pooling together the results of different models into a meaningless summary. For illustration, we did an additional MCMC run for the model with one location factor (15) and two dispersion factors (1 and 8), applying a software option that bypasses the model-selection steps and forces these factors to stay active. The results are summarized in Table 3. The quantities LCL and UCL provide 95% credibility intervals based on the quantiles of the empirical distributions. Information on β_{15} is not available in this way because the parameter has been integrated out.

It is interesting to compere the results just obtained with an analysis based on all 28 runs in Table 2. Results of the same models as in Figure 2

Table 3: Summary of key parameters when one location effect and two dispersion effects are forced in the model.

Parameter	Median	Mean	Std. deviation	LCL	UCL
σ_γ	1.83	1.92	0.76	0.71	3.39
γ_1	−1.42	−1.42	0.60	−2.64	−0.26
γ_8	1.06	1.07	0.58	−0.07	2.24

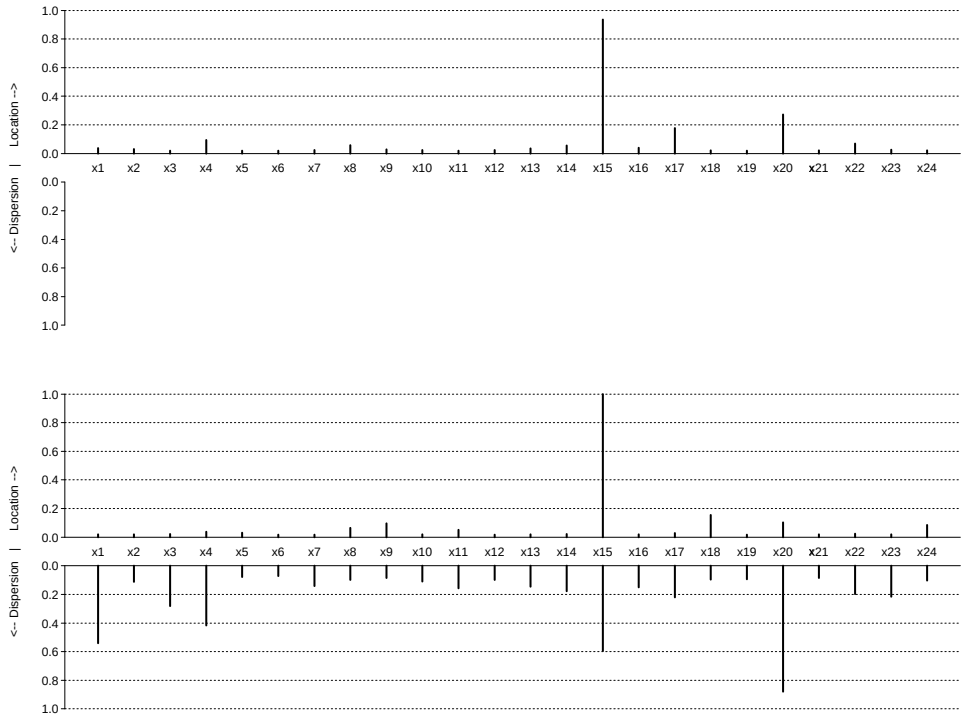

Figure 4: Posterior probability plots for the complete 28-run design (Williams, 1968), using the same models as in Figure 2.

are shown in Figure 4. For the location-only model, only factor 15 has a high posterior probability. As before, adding dispersion effects to the model (the bottom panel of Figure 4) strengthens factor 15's importance as a location predictor, and weakens the others. Factor 1 stands out again as a dispersion predictor, along with three or four others not deemed important in the analysis of the smaller data set.

5 Conclusions

The purpose of this article has been to investigate an adaptation of the effect-sparsity model of Box and Meyer (1986a) for a more general situation where there are both location and dispersion effects. The problem is made tractable by means of MCMC methods, particularly the reversible-jump algorithm of Green (1995).

In the instance of small, unreplicated experiments, the emphasis is less on obtaining a definitive analysis than on obtaining guidance for further investigation. Surely we do not pretend that a definitive analysis can be obtained in a situation where there are many more parameters than observations, such as the supersaturated experiment described above. But the MCMC results can be usefully viewed as a collection of trial analyses, and as such can be quite useful for guiding one's decisions in designing a subsequent experiment.

Clearly, there are generalizations of this work worth investigating. One limitation of the methods set out here is that there are no restrictions on the models considered. This is not a problem with the examples we have shown, but it is often the case that there is a hierarchy in the predictors, such as

interactions or polynomials, where some predictors should not be included unless others are also included. Moreover, we might want the values or priors for α and ϕ to depend on the order of the predictor. Such scenarios for the location model are considered in Hamada and Wu (1992); Chipman (1996); Chipman et al. (1997) but need to be extended to the dispersion model. Simply imposing constraints on predictor selection in the MCMC algorithm is theoretically valid, but it leads to very poor mixing behavior in practice, especially with many predictors.

Another area of importance is complex designs such as are discussed in Wolfinger and Tobias (1998) that incorporate random effects or blocking, such as split-plot designs. Dong and Nair (1999) considers the Bayesian analysis of location and dispersion in a few such designs, where the design is balanced. The model is similar in some ways to the one here, but it avoids the reversible-jump aspect by using mixtures of two normals in place of (6) and (7).

A Posterior pdf

Let $\boldsymbol{\theta}$ denote the vector of all parameters, and let $\pi(\alpha, \phi, c, d)$ embody any prior densities that we place on α, ϕ, c, and d. The joint posterior of all of the parameters may be found by taking the product of $\pi(\alpha, \phi, c, d)$ and the densities specified in equations (4) through (9). It is made more tractable by working in terms of the Bernoulli random variables $\boldsymbol{\delta}$, $\boldsymbol{\omega}$ (see (10), (11)) and the conditional distributions of $\boldsymbol{\beta}|\boldsymbol{\delta}$ and $\boldsymbol{\gamma}|\boldsymbol{\omega}$. We thus obtain the posterior

density

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{y}) \;\; &\propto \;\; \pi(\alpha,\phi,c,d)\cdot\frac{\alpha^{m_\beta}(1-\alpha)^{p-m_\beta}\phi^{m_\gamma}(1-\phi)^{q-m_\gamma}\det(\boldsymbol{W})^{1/2}}{(2\pi)^{(m_\beta+m_\gamma)/2}\lambda c^{m_\beta}d\sigma_\epsilon^{n+m_\beta+2\nu+2}\sigma_\gamma^{m_\gamma}} \\ &\quad\times\exp\left\{-\frac{\boldsymbol{\beta}'\boldsymbol{\beta}}{2c^2\sigma_\epsilon^2}-\frac{\beta_0^2}{2d^2\sigma_\epsilon^2}-\frac{\boldsymbol{\gamma}'\boldsymbol{\gamma}}{2\sigma_\gamma^2}-\frac{1}{\psi\sigma_\epsilon^2}\right\} \\ &\quad\times\exp\left\{-\frac{(\boldsymbol{y}-\beta_0\mathbf{1}-\boldsymbol{X}_{\boldsymbol{\delta}}\boldsymbol{\beta}_{\boldsymbol{\delta}})'\boldsymbol{W}(\boldsymbol{y}-\beta_0\mathbf{1}-\boldsymbol{X}_{\boldsymbol{\delta}}\boldsymbol{\beta}_{\boldsymbol{\delta}})}{2\sigma_\epsilon^2}\right\} \qquad (19)\end{aligned}$$

where $m_\beta = \mathbf{1}'\boldsymbol{\delta}$, $m_\gamma = \mathbf{1}'\boldsymbol{\omega}$, $\boldsymbol{X}_{\boldsymbol{\delta}}$ is the $n \times m_\beta$ matrix having columns $\{\boldsymbol{x}_j : \delta_j = 1\}$, and $\boldsymbol{\beta}_{\boldsymbol{\delta}}$ has elements $\{\beta_j : \delta_j = 1\}$. Let

$$\begin{aligned}\tilde{\boldsymbol{X}} &= \left[\begin{array}{c|c} \mathbf{1} & \boldsymbol{X}_{\boldsymbol{\delta}} \\ \hline d^{-2} & 0 \\ \hline \mathbf{0} & c^{-2}\boldsymbol{I}\end{array}\right], \\ \tilde{\boldsymbol{y}} &= [\,\boldsymbol{y}\mid 0\mid \mathbf{0}\,]', \\ \tilde{\boldsymbol{\beta}} &= [\,\beta_0\mid \boldsymbol{\beta}_{\boldsymbol{\delta}}\,]' \\ \tilde{\boldsymbol{W}} &= \mathrm{diag}(\boldsymbol{W},1,\boldsymbol{I}) \qquad (20)\end{aligned}$$

where the partitioning of $\tilde{\boldsymbol{X}}$ has $(n,1,m_\beta)$ rows and $(1,m_\beta)$ columns, and the other structures conform. Also note that, due to the requirement that $\boldsymbol{Z}$ be centered, $\det(\boldsymbol{W}) = 1$. Thus, (19) simplifies to

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{y}) \;\; &\propto \;\; \pi(\alpha,\phi,c,d)\cdot\frac{\alpha^{m_\beta}(1-\alpha)^{p-m_\beta}\phi^{m_\gamma}(1-\phi)^{q-m_\gamma}}{(2\pi)^{(m_\beta+m_\gamma)/2}\lambda c^{m_\beta}d\sigma_\epsilon^{n+m_\beta+2\nu+2}\sigma_\gamma^{m_\gamma}}\cdot\exp\left\{-\frac{\boldsymbol{\gamma}'\boldsymbol{\gamma}}{2\sigma_\gamma^2}\right\} \\ &\quad\times\exp\left\{-\frac{(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})'\tilde{\boldsymbol{W}}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}})}{2\sigma_\epsilon^2}-\frac{1}{\psi\sigma_\epsilon^2}\right\} \qquad (21)\end{aligned}$$

The last factor can be integrated with respect to $\tilde{\boldsymbol{\beta}}$, yielding

$$\frac{(2\pi)^{(m_\beta+1)/2}\sigma_\epsilon^{m_\beta+1}}{\sqrt{\det(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{W}}\tilde{\boldsymbol{X}})}}\exp\left\{-\frac{2/\psi+S(\boldsymbol{\delta},\boldsymbol{\gamma})}{2\sigma_\epsilon^2}\right\} \qquad (22)$$

where $S\boldsymbol{\delta}, \boldsymbol{\gamma})$ is defined in (12). Due to the constraints on $\boldsymbol{Z}$, $\det(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{W}}\tilde{\boldsymbol{X}}) = \det(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})$. Substituting (22) for the last factor in (21), then integrating with respect to σ_ϵ, we obtain

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{y}) \quad &\propto \quad \pi(\alpha,\phi,c,d)\cdot\frac{\alpha^{m_\beta}(1-\alpha)^{p-m_\beta}\phi^{m_\gamma}(1-\phi)^{q-m_\gamma}}{(2\pi)^{m_\gamma/2}\lambda c^{m_\beta}d\sigma_\gamma^{m_\gamma}\sqrt{\det(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})}} \\ &\quad \times\exp\left\{-\frac{\boldsymbol{\gamma}'\boldsymbol{\gamma}}{2\sigma_\gamma^2}\right\}\{2/\psi+S(\boldsymbol{\delta},\boldsymbol{\gamma})\}^{-(\nu+n/2)} \qquad (23)\end{aligned}$$

Similar results are given in George and McCulloch (1997) and Chipman (1998) for the case where there is no dispersion model, i.e., $\boldsymbol{W} = \boldsymbol{I}$.

B MCMC Samplers

Full-conditional distributions and other entities required by the MCMC methods are often easily obtained by inspection of (23), leaving out any factors that do not involve the parameter in question. Thus, the numerator and denominator of (13) is obtained by substituting $\delta_j = 1$ and $\delta_j = 0$, respectively, in (23). The quantities in (14) and (18) are a straightforward application of the Metropolis-Hastings algorithm. Care must be used in including quantities that depend on parameters not explicitly shown; for example, m_β depends on $\boldsymbol{\delta}$, $\tilde{\boldsymbol{X}}$ depends on c, d, and $\boldsymbol{\delta}$, and $\tilde{\boldsymbol{W}}$ depends on $\boldsymbol{\omega}$ and $\boldsymbol{\gamma}$.

The reversible-jump method in Algorithm ARC requires somewhat more explanation. It is closely related to the Metropolis-Hastings algorithm, but additional factors must be included to ensure that the detailed-balance property of Markov chains will be satisfied when transitions are made between parameter spaces of different dimension. Following Green (1995) to derive the quantity A_{+j} used in Move A, we need to include the factor P_R to adjust

for the fact that the reverse jump is Move R, which is considered with probability P_R. Also, Move A involves a $N(0, \sigma_\gamma^2)$ "proposal" distribution. The mapping of the remaining $\gamma_l, l \neq j$ is trivial, so the Jacobian is 1. Thus,

$$A_{+j} = \frac{\pi(\boldsymbol{\theta}|\boldsymbol{y}, \gamma_j = 1)}{\pi(\boldsymbol{\theta}|\boldsymbol{y}, \gamma_j = 0)} \cdot P_R \cdot \exp\{-\tfrac{1}{2}\gamma_j^2/\sigma_\gamma^2\} \cdot 1 \tag{24}$$

The exponential factor cancels with a like factor in the ratio of the posterior densities, and we obtain the result in (16).

Move C does not involve a change of dimension, and thus (17) is an ordinary Metropolis-Hastings step with a symmetric proposal distribution (i.e., a Metropolis step).

References

Bergman, B. and Hynén, A. (1997), "Dispersion Effects From Unreplicated Designs in the 2^{k-p} Series," *Technometrics*, 39, 191–198.

Box, G. E. P. and Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, Wiley, New York.

Box, G. E. P. and Meyer, R. D. (1986a), "An Analysis for Unreplicated Fractional Factorials," *Technometrics*, 28, 11–18.

Box, G. E. P. and Meyer, R. D. (1986b), "Dispersion Effects From Fractional Designs," *Technometrics*, 28, 19–27, Correction: vol. 29, p. 250.

Brooks, S. P. and Giudici, P. (1999), "Diagnosing Convergence of Reversible Jump MCMC Algorithms," in *Bayesian Statistics VI*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford University Press.

Brooks, S. P. and Roberts, G. O. (1999), "Diagnosing Convergence of Markov Chain Monte Carlo Algorithms," *Statistics in Computing*, 8, 319–335.

Carlin, B. P. and Chib, S. (1995), "Bayesian Model Choice via Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society, Series B*, 57, 473–484.

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

Chipman, H. (1998), "Fast Model Search for Designed Experiments," in *Proceedings of the Conference for Quality Improvement Through Statistical Methods*, Birkhauser, to appear.

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.

Cowles, M. K. and Carlin, B. P. (1996), "Markov Chain Monte Carlo Convergence Diagnostics: a Comparative Review," *Journal of the American Statistical Association*, 91, 883–904.

Dong, X. and Nair, V. (1999), "Bayesian Analysis of the Location-Dispersion Model," Tech. rep., University of Michigan.

Engel, J. and Huele, A. F. (1996), "A Generalized Linear Modeling Approach to Robust Design," *Technometrics*, 38, 365–373.

Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Geman, S. and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

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

George, E. I. and McCulloch, R. E. (1997), "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7, 339–373.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.) (1996), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Gill, P. E., Golub, G. H., Murray, W., and Saunders, M. A. (1974), "Methods for Modifying Matrix Factorizations," *Mathematics of Computation*, 28, 503–536.

Gill, P. E., Murray, W., and Saunders, M. A. (1975), "Methods for Computing and Modifying the LDV Factors of a Matrix," *Mathematics of Computation*, 29, 1051–1077.

Green, P. J. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711–732.

Haaland, P. D. and O'Connell, M. R. (1995), "Inference for Effect-Saturated Fractional Factorials," *Technometrics*, 37, 82–93.

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

Hastings, W. K. (1970), "Monte Carlo Sampling Methods using Markov Chains and Their Applications," *Biometrika*, 57, 97–109.

Lin, D. K. J. (1993), "A new class of supersaturated designs," *Technometrics*, 35, 28–31.

Mays, D. P. and Myers, R. H. (1996), "Bayesian Approach for the Design and Analysis of a Two Level Factorial Experiment in the Presence of Dispersion Effects," *Communications in Statistics, Part A—Theory and Methods*, 25, 1409–1428.

Mengersen, K. L., Robert, C. P., and Guihenneuc-Jouyaux, C. (1998), "MCMC Convergence Diagnostics: A 'RevieWWW'," in *Bayesian Statistics VI*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford University Press.

Nelder, J. A. and Lee, Y. (1991), "Generalized Linear Models for the Analysis of Taguchi-Type Experiments," *Applied Stochastic Models and Data Analysis*, 7, 107–120.

Nelder, J. A. and Lee, Y. (1998), "Joint Modelling of Mean and Dispersion," *Technometrics*, 40, 168–175, Letter, with responses from other authors.

Taguchi, G. and Wu, Y. (1986), *Introduction to Off-Line Quality Control*, Central Japan Quality Control Association, Nagoya, Japan.

Williams, K. R. (1968), "Designed Experiments," *Rubber Age*, 100, 65–71.

Wolfinger, R. D. and Tobias, R. D. (1998), "Joint Estimation of Location, Dispersion, and Random Effects in Robust Design," *Technometrics*, 40, 62–71.