Exchange Algorithms for Constructing Model-Robust Experimental Designs

Byran J. Smucker[1]*, Enrique del Castillo[2]†, and James L. Rosenberger[3]‡

[1]Department of Statistics

Miami University, Oxford, OH 45056

[2]Department of Industrial & Manufacturing Engineering

The Pennsylvania State University, University Park, PA 16802

[3]Department of Statistics

The Pennsylvania State University, University Park, PA 16802

Abstract

Optimal experimental design procedures, utilizing criteria such as $\mathcal{D}$-optimality, are useful for producing designs for quantitative responses, often under nonstandard conditions such as constrained design spaces. However, these methods require *a priori* knowledge of the exact form of the response function, an often unrealistic assumption. Model-robust designs are those which, from our perspective, are efficient with respect to a set of possible models. In this paper, we develop a model-robust technique motivated

*Dr. Smucker is an Assistant Professor of Statistics at Miami University in Oxford, OH and a Student Member of ASQ. His e-mail address is absmucker@gmail.com.

†Dr. del Castillo is a Distinguished Professor of Industrial and Manufacturing Engineering and a Professor of Statistics at the Pennsylvania State University. His e-mail address is exd13@psu.edu.

‡Dr. Rosenberger is a Professor of Statistics at The Pennsylvania State University. His e-mail address is jlr@psu.edu.

by a connection to multiresponse $\mathcal{D}$-optimal design. This link spawns a generalization of the modified Fedorov exchange algorithm, which is then used to construct exact model-robust designs. We also study the effectiveness of designs robust for a small set of models compared to designs which account for much larger sets. We give several examples and compare our designs with two model-robust procedures in the literature.

Keywords: $\mathcal{D}$-optimality, multiresponse design, robust design, model space

Introduction and Motivation

Since Kiefer (1959) debuted the idea of optimal design of experiments, a vast literature has grown up around the notion of choosing a design based upon some numerical criterion. The most common is $\mathcal{D}$-optimality, which chooses the design minimizing the generalized variance of the regression parameter estimates. Though standard designs can be used in most design situations, optimal procedures are useful when, for instance, there are constraints on the design space or some factors are categorical. However, optimal design procedures have been criticized (Box and Draper 1959) because they require complete knowledge of the form of the regression function, though this knowledge is rarely at hand. Subsequently, techniques have been developed which produce designs that are in some way robust to departures from the assumed model.

For instance, optimal designs are often used in mixture experiments because of the constrained nature of the design region. Heinsman and Montgomery (1995) describe an experiment involving four surfactant mixture factors, for the purpose of optimizing a household product. There were further factor restrictions beyond the mixture constraint, which made optimal design a natural choice. However, such a design would require the complete specification of the form of the mixture regression model. For instance, a special cubic Sheffé polynomial model might be assumed, though it is unknown before the experiment whether this is the correct model. We provide a procedure which allows the experimenter to obtain a design which does not assume a single model form, but rather accounts for a class of user-specified models. We revisit this example later.

Model-robustness has enjoyed significant development over the years, primarily in the hands of theoreticians whose work has provided insights into specific problems and the tradeoff between bias and variance in the assessment of optimal designs; see, for instance Montepiedra and Fedorov (1997), Dette and Franke (2001), Fang and Wiens (2003), Zhou (2008). On a practical level, much of this work lacks an intuitive framework within which an experimenter might work. In fact, as Chang and Notz (1996) point out in a review of similar—though earlier—work, these model-robust methods have more value for warning of the dangers of ignoring the issue than for providing designs that can be adopted by practitioners.

Even further, nearly all of this research employs Kiefer's continuous design theory which, while mathematically elegant and tractable, produces designs optimal for asymptotically large run sizes. In contrast, most applications in the physical sciences and engineering require optimal designs for n runs, where n is a relatively small number. These designs are called discrete, or exact, and are denoted by ξ_n (in the next section, we refer briefly to asymptotic designs, and denote them by ξ). Consequently, commercial software implementations employ exchange algorithms for fixed sample sizes, including the Fedorov exchange algorithm (Fedorov 1972), DETMAX (Mitchell 1974), and the k-exchange algorithm (Johnson and Nachtsheim 1983).

There is remarkably little work done in accessible methods for exact (small-sample) model-robust designs. A mean squared error criterion reminiscent of Box and Draper (1959) was proposed by Welch (1983), along with a DETMAX-like exchange algorithm, and Fang and Wiens (2000) give a similar approach, using a minimax average mean squared error criterion utilizing simulated annealing. DuMouchel and Jones (1994) use a Bayesian approach to provide some protection against specified terms not in the assumed model, but their method requires specification of a prior precision parameter and does not explicitly guard against more than two models; i.e. the assumed model and one that includes the potential terms. Still, this approach formalizes the *ad hoc* practice of adding center points to test for lack of fit and has spawned significant follow-up work, such as Neff (1996), Goos

et al. (2005), and Jones et al. (2008). Heredia-Langner et al. (2004) allow protection against multiple models by utilizing a desirability function to incorporate information about each possible model. The necessary optimization is performed using a genetic algorithm, which introduces additional complexity in implementation.

We propose a new method which produces exact designs robust for a set of user-defined possible models. These ideas are motivated by a connection between multiresponse regression (Zellner 1962), multiresponse optimal design (Fedorov 1972), and a continuous model-robust optimal design technique due to Läuter (1974). To implement these ideas, we develop a model-robust exchange algorithm which generalizes existing univariate methods.

The paper is organized as follows. In the next section we give the technical background and describe the basic approach taken to find model-robust designs. We then review some basic univariate exchange algorithms and give a generalization which is used to find model-robust designs. We next give several examples illustrating our method and compare our designs to those of DuMouchel and Jones (1994) and Heredia-Langner et al. (2004). Following this, we study the effect of designing with respect to just a small fraction of all possible models, and conclude with discussion of the procedure and its results.

Setting and Proposed Approach

Suppose one is interested in performing an experiment with a single quantitative response variable, y, and a factors (quantitative or categorical), $\mathbf{x} = (x_1, \ldots, x_a)$. We assume that the classical univariate linear regression model will be fit, where $y_i = \mathbf{f}'(\mathbf{x}_i)\boldsymbol{\beta} + \epsilon_i, i = 1, \ldots, n$ with $\boldsymbol{\beta}$ a p-vector of parameters and $\mathbf{f}(\mathbf{x})$ the p-vector valued model function, though p and the precise form of $\mathbf{f}(\mathbf{x})$ are unknown (for convenience, we often write $\mathbf{f}$, suppressing the argument). In matrix notation, we have $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{y}$ is an n-vector, $\mathbf{X}$ is an $n \times p$ expanded design matrix, and $\boldsymbol{\epsilon}$ is also an n-vector with $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $Var(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}_n$. We assume also that the least squares criterion is used to estimate $\boldsymbol{\beta}$, in which case the estimator is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ with $Cov(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

To fit such a model, the design must be chosen and y_i observed at each of the design points, $\mathbf{x}_i$. Let χ be the design space, Ξ be the set of all possible designs and $\xi_n \in \Xi$ be a discrete, n-point design:

$$\xi_n = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{pmatrix}. \tag{1}$$

We define the information matrix in this case as $\mathbf{M}(\xi_n) = \sigma^{-2} \sum_{i=1}^{n} \mathbf{f}(\mathbf{x}_i)\mathbf{f}'(\mathbf{x}_i) = (\mathbf{X}'\mathbf{X})/\sigma^2 = [Cov(\hat{\boldsymbol{\beta}})]^{-1}$.

An optimal design approach would attempt to find the n points, $\mathbf{x}_i \in \chi, i = 1, \dots, n$, such that some criterion, $\phi(\mathbf{M}(\xi_n))$, is optimized. Many criteria have been proposed, but perhaps the most popular and mathematically tractable is the $\mathcal{D}$-optimality, for which $\phi(\mathbf{M}(\xi_n)) = |\mathbf{M}(\xi_n)|$. Such an optimal design minimizes the volume of the confidence ellipsoid of the parameters.

Since the precise form of $\mathbf{f}(\mathbf{x})$ is generally not known, we make the weaker assumption that there exists a set of r possible models $\mathcal{F}$ that might be fit. Läuter (1974) presented this idea for asymptotic designs ξ, and introduced a model-robust criterion similar to $\phi\left(\mathbf{M}_{\mathcal{F}}(\xi)\right) = \prod_{\mathbf{f}\in\mathcal{F}} |\mathbf{M}_{\mathbf{f}}(\xi)|$, where $\mathbf{M}_{\mathcal{F}}(\xi) = (\mathbf{M}_{\mathbf{f}_1}(\xi), \dots, \mathbf{M}_{\mathbf{f}_r}(\xi))$ and $\mathbf{M}_{\mathbf{f}}(\xi)$ is the information matrix for model $\mathbf{f}$. Thus, the design which maximizes $\phi(\mathbf{M}_{\mathcal{F}}(\xi))$ over all possible designs can be considered robust to the models in $\mathcal{F}$. Cook and Nachtsheim (1982) utilized this idea to develop linear-optimal designs focusing on prediction. Later, Dette (1990) used the theory of canonical moments to give more explicit solutions for this product criterion. These papers, however, are limited to continuous designs and unconstrained cuboidal design regions. Other authors have used this idea for exact designs, including Li and Nachtsheim (2000) who concentrated on factorial designs for which a prespecified number of two-factor interactions were of interest, utilizing work by Sun (1993). Jones et al. (2009) also extended the same basic idea to supersaturated designs.

Our discrete approach springs from Laüter's idea, since allowing the experimenter to define a class of possible models is practically compelling. When model-robustness is viewed in this way, it is closely related to multiresponse optimal design; see Fedorov (1972), Khuri

and Cornell (1996), Chang (1997), and Atashgah and Seifi (2009). These methods are based upon a multiresponse regression model due to Zellner (1962) which allows the functional form of the factors to be different for each response and can produce more precise estimates of the regression parameters by considering the covariance structure of the responses. Thus, a multiresponse optimal design is one which is optimal for all the responses together, which is essentially the goal of the univariate model-robust criterion of Läuter (1974), when "responses" are replaced by "models". After we had begun this work, we discovered a technical report (Emmett et al. 2007) which makes the same connection, though the basis of our work is independent of theirs.

Consequently, one interpretation of the model-robust designs using Laüter's criterion with r possible models is that they are also $\mathcal{D}$-optimal for the multiresponse model with r uncorrelated responses, provided that each of the r responses has a functional form that corresponds one-to-one with the r possible models. This is because the multiresponse $\mathcal{D}$-optimal criterion simplifies to the Laüter's model-robust product criterion. Furthermore, there are multiresponse results (Bischoff 1993, Kurotschka and Schwabe 1996) which prove that when the response models are nested (where nested implies that when the response models are ordered by their number of parameters, from smallest to largest, each successive model includes the previous model as a subset), the multiresponse $\mathcal{D}$-optimal designs are invariant to the covariance structure of the responses. This suggests that when the set of models are nested, the model-robust design is the same as the associated multiresponse $\mathcal{D}$-optimal design, even if the responses are correlated.

A Model-Robust Exchange Algorithm

In this section we first review the basic univariate exchange algorithms upon which our methods are based, and then present a generalization to the matrix-updating formulas used in the univariate procedures. Finally, we introduce our model-robust exchange algorithm, which utilizes this generalization to avoid calculating determinants when evaluating poten-

tial exchanges.

Univariate Exchange Algorithms

The first univariate exchange algorithm (Fedorov 1972) considered exchanges between each design point and points in a candidate list, a discretized version of the design space. At each iteration, the exchange was made which most increases the determinant of the information matrix. He exploited a determinant-updating formula to alleviate the considerable computational burden this problem imposed. Specifically, given design ξ_n and model $\mathbf{f}$, he showed that if $\mathbf{x}_j \in \xi_n$ is exchanged for $\mathbf{x} \in \chi$ resulting in the new design $\tilde{\xi}_n$,

$$|\mathbf{M}_{\mathbf{f}}(\tilde{\xi}_n)| = |\mathbf{M}_{\mathbf{f}}(\xi_n)| \left(1 + \Delta_{\mathbf{f}}(\mathbf{x}_j, \mathbf{x}, \xi_n)\right) \tag{2}$$

where

$$\Delta_{\mathbf{f}}(\mathbf{x}_j, \mathbf{x}, \xi_n) = \mathbf{V}_{\mathbf{f}}(\mathbf{x}, \xi_n) - \mathbf{V}_{\mathbf{f}}(\mathbf{x}, \xi_n)\mathbf{V}_{\mathbf{f}}(\mathbf{x}_i, \xi_n) + \mathbf{V}^2_{\mathbf{f}}(\mathbf{x}, \mathbf{x}_j, \xi_n) - \mathbf{V}_{\mathbf{f}}(\mathbf{x}_j, \xi_n) \tag{3}$$

under the assumption that $\sigma^2 = 1$, with $\mathbf{V}_{\mathbf{f}}(\mathbf{x}, \xi_n) = \mathbf{f}'(\mathbf{x})\mathbf{M}^{-1}_{\mathbf{f}}(\xi_n)\mathbf{f}(\mathbf{x})$ and $\mathbf{V}_{\mathbf{f}}(\mathbf{x}, \mathbf{x}_j, \xi_n) = \mathbf{f}'(\mathbf{x})\mathbf{M}^{-1}_{\mathbf{f}}(\xi_n)\mathbf{f}(\mathbf{x}_j)$. We can also update the inverse of the information matrix using Lemma 3.3.1 in Fedorov (1972), using notation from Meyer and Nachtsheim (1995):

$$\mathbf{M}^{-1}_{\mathbf{f}}(\tilde{\xi}_n) = \mathbf{M}^{-1}_{\mathbf{f}}(\xi_n) - \mathbf{M}^{-1}_{\mathbf{f}}(\xi_n)\mathbf{F}_1(\mathbf{I}_2 + \mathbf{F}_2'\mathbf{M}^{-1}_{\mathbf{f}}(\xi_n)\mathbf{F}_1)^{-1}\mathbf{F}_2'\mathbf{M}^{-1}_{\mathbf{f}}(\xi_n) \tag{4}$$

with $\mathbf{F}_1 = [\mathbf{f}(\mathbf{x}), -\mathbf{f}(\mathbf{x}_j)]$ and $\mathbf{F}_2 = [\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}_j)]$. The Fedorov algorithm is as follows:

1. Initialize algorithm: Begin with a nonsingular design ξ_n; construct grid, $\mathcal{G} \subset \chi$
2. Let $j = 1$.
3. For design point $\mathbf{x}_j$, calculate $\Delta_{\mathbf{f}}(\mathbf{x}_j, \mathbf{x}, \xi_n)$ as in (3) for all $\mathbf{x} \in \mathcal{G}$. Choose $\mathbf{x}^*_j = \arg\max_{\mathbf{x}\in\chi} \Delta_{\mathbf{f}}(\mathbf{x}_j, \mathbf{x}, \xi_n)$.
4. Increment j and if $j < n$ return to Step 3. Else choose $j^* = \arg\max_{j\in\{1,\cdots,n\}} \Delta_{\mathbf{f}}(\mathbf{x}_j, \mathbf{x}^*_j, \xi_n)$

and exchange $\mathbf{x}_{j^*}$ and $\mathbf{x}^*_{j^*}$, resulting in a new design $\tilde{\xi}_n$. Update the determinant via (2).

5. Update the inverse of the information matrix via (4).

6. If $\Delta_{\mathbf{f}}(\mathbf{x}_{j^*}, \mathbf{x}^*_{j^*}, \xi_n) < \epsilon$, STOP. Else set $\xi_n = \tilde{\xi}_n$ and return to Step 2.

This algorithm generates a convergent nondecreasing sequence of determinants, but will not in general converge to the global optimum. Therefore, it is necessary to run many instances of the algorithm each with a randomly generated initial design. Despite the cheap determinant updates, the primary drawback to Fedorov's algorithm is its computational demands since n optimizations are required during each iteration.

Cook and Nachtsheim (1980) proposed a modified Fedorov exchange algorithm, which mimics Fedorov's original procedure but exchanges each $\mathbf{x}_j$ and $\mathbf{x}^*_j$ in Step 3. This capitalizes on each of the n optimizations that are performed during each iteration, and seems to be as effective as its archetype. It is actually a special case of the k-exchange algorithm (Johnson and Nachtsheim 1983), which considers only the k least critical design points (those with the smallest prediction variance) for exchange.

In the remainder of this paper, we develop a multiresponse generalization of the modified Fedorov exchange algorithm and use it to construct single response model-robust designs. We focus on this algorithm since we found it to be faster than the original Fedorov algorithm while producing better designs than the k-exchange. Similar extensions to other existing univariate algorithms, such as DETMAX (Mitchell 1974), BLKL (Atkinson et al. 2007), and coordinate-exchange (Meyer and Nachtsheim 1995), could be developed. The latter does not require a candidate list and is computationally attractive, but it does not accommodate constraints on the design space. A more recent algorithm (Piepel et al. 2005) suggests this limitation might be overcome, but personal communication with the first and last authors of this paper indicates that for experiments with a mixture constraint, in particular, further improvements are necessary before its general use can be recommended.

Model-Robust Modified Fedorov Exchange Algorithm

As in Läuter (1974) we consider, instead of a single model, a finite set of r models $\mathcal{F}$ from which the experimenter believes the true model form can be chosen. More specifically, let ξ_n be an n-point design and $\mathbf{M}_{\mathbf{f}_i}(\xi_n)$ be the information matrix for model i where $\mathbf{f}_i \in \mathcal{F}$, $i = 1, \cdots, r$. Suppose that we exchange a design point $\mathbf{x}_j$ for an arbitrary point $\mathbf{x}$ in the design region, resulting in a new design $\tilde{\xi}_n$. Then the model-robust optimization criteria can be written as:

$$\begin{aligned} \boldsymbol{\phi}(\mathbf{M}_{\mathcal{F}}(\tilde{\xi}_n)) &= \prod_{i=1}^{r} \left| \mathbf{M}_{\mathbf{f}_i}(\tilde{\xi}_n) \right| \\ &= \prod_{i=1}^{r} \left| \mathbf{M}_{\mathbf{f}_i}(\xi_n) \right| \left(1 + \Delta_{\mathbf{f}_i}(\mathbf{x}_j, \mathbf{x}, \xi_n)\right) \\ &= \boldsymbol{\phi}(\mathbf{M}_{\mathcal{F}}(\xi_n)) \prod_{i=1}^{r} (1 + \Delta_{\mathbf{f}_i}(\mathbf{x}_j, \mathbf{x}, \xi_n)) \end{aligned} \tag{5}$$

so that for each iteration of the algorithm, we need to just calculate and maximize $\prod_{i=1}^{r}(1 + \Delta_{\mathbf{f}_i}(\mathbf{x}_j, \mathbf{x}, \xi_n))$ where $\Delta_{\mathbf{f}_i}(\mathbf{x}_j, \mathbf{x}, \xi_n)$ is calculated as in (3) for model $\mathbf{f}_i$. We make a slight adjustment to this criterion so our algorithm will not choose to exchange a point that is so bad that $(1 + \Delta_{\mathbf{f}_i}(\mathbf{x}_j, \mathbf{x}, \xi_n)) < 0$ for an even number of models, which would result in a positive value of our criterion even though the exchange is undesirable. Thus, we choose the exchange which maximizes

$$\prod_{i=1}^{r} (1 + \Delta_{\mathbf{f}_i}(\mathbf{x}_j, \mathbf{x}, \xi_n)) \mathbb{I}(1 + \Delta_{\mathbf{f}_i}(\mathbf{x}_j, \mathbf{x}, \xi_n) > 0) \tag{6}$$

where $\mathbb{I}$ is the indicator function.

Based on the above development, the algorithm is as follows:

1. Initialize algorithm: Begin with a nonsingular design ξ_n; construct grid, $\mathcal{G} \subset \chi$.

2. Let $j = 1$.

3. For design point $\mathbf{x}_j$, calculate (6) for all $\mathbf{x} \in \mathcal{G}$. Choose $\mathbf{x}_j^* = \arg\max_{\mathbf{x} \in \chi} \prod_{i=1}^{r} (1 +$

$\Delta_{\mathbf{f}_i}(\mathbf{x}_j, \mathbf{x}, \xi_n))\mathbb{I}(1 + \Delta_{\mathbf{f}_i}(\mathbf{x}_j, \mathbf{x}, \xi_n) > 0)$.

4. Perform exchange $\mathbf{x}_j^*$ for $\mathbf{x}_j$, and call the updated design $\tilde{\xi}_n$. Calculate $\left|\mathbf{M}_{\mathbf{f}_i}(\tilde{\xi}_n)\right|$ and $\mathbf{M}_{\mathbf{f}_i}^{-1}(\tilde{\xi}_n)$ for each model via (2) and (4), respectively.

5. Increment j and if $j < N$ return to Step 3. Else, if $\max_j \prod_{i=1}^{r}(1 + \Delta_{\mathbf{f}_i}(\mathbf{x}_j, \mathbf{x}_j^*, \xi_n)) < 1 + \epsilon$, STOP. Else set $\xi_n = \tilde{\xi}_n$ and return to Step 2.

As in the standard exchange algorithms, to find a global optimum for larger problems it is necessary to perform many runs of the algorithm using different initial designs.

Examples

In this section we present several examples illustrating the proposed *model-robust modified Fedorov* (MRMF) exchange algorithm, and compare it with two other exact model-robust design methods in the literature. Before giving the examples, we will briefly describe these methods and discuss how the designs will be evaluated.

DuMouchel and Jones (1994) use a Bayesian approach to provide protection against higher-order terms. They set s_1 terms as primary and s_2 terms as potential and after scaling the two groups to make them nearly orthogonal, they assume an informative prior for the potential terms and calculate a posterior distribution for the parameters with variance $\mathbf{A} = [\mathbf{X}'\mathbf{X} + \mathbf{K}/\tau^2]^{-1}$, where $\mathbf{X} = (\mathbf{X}_{pri}|\mathbf{X}_{pot})$ and $\mathbf{K}$ is a $(s_1 + s_2) \times (s_1 + s_2)$ diagonal matrix with 0 on the first s_1 diagonals and 1 on the last s_2. The prior variance parameter, τ, is to be chosen by the user. Once they have this posterior variance, they simply choose the design that minimizes $|\mathbf{A}|$ using slightly adjusted exchange algorithms.

A distinct advantage of this method is that it can provide protection against models with more parameters than observations. On the other hand, it is not designed to produce model-robust designs with respect to more than two models. Since it is a prominent and rare model-robust technique for exact designs, we compare its results to ours. Difficulties associated with this method are the choice of the prior precision value, $\frac{1}{\tau}$, and how to

designate the primary and potential terms. We use $\frac{1}{\tau} = 1$, as recommended by DuMouchel and Jones, but also include designs based upon $\frac{1}{\tau} = 16$. Because of the structure of $\mathbf{A}$, larger prior precision values will result in less consideration of the potential terms as manifested by lower efficiencies for models involving those terms. We also generally assume more primary terms as opposed to less. The results are based upon the implementation of this method in the $SAS^{®}$ software's PROC OPTEX (SAS 2004).

Heredia-Langner et al. (2004) used a genetic algorithm to calculate exact model-robust designs. They consider r possible models and use a genetic algorithm to optimize a desirability function which incorporates the determinants of the information matrices of each of the models. Their procedure does not require a candidate list, though implementation of a tuned genetic algorithm is not trivial. Examples 1 and 3 are taken from their paper, which allows comparisons to be made.

We compare designs on the basis of efficiencies with respect to each model $\mathbf{f} \in \mathcal{F}$. The $\mathcal{D}$-efficiency for model $\mathbf{f}$ is $E_{\mathbf{f}}(\xi_n) = \left(\frac{|\mathbf{M}_{\mathbf{f}}(\xi_n)|}{|\mathbf{M}_{\mathbf{f}}(\xi^*_{n,\mathbf{f}})|}\right)^{1/p}$ where $\xi^*_{n,\mathbf{f}}$ is the design optimal for $\mathbf{f}$ alone, and p is the number of parameters for model $\mathbf{f}$. Note that we express the $\mathcal{D}$-efficiency as $E_{\mathbf{f}}$ in the tables below, since the appropriate design is evident. Since determinants can roughly be viewed as measures of volume, this quantity takes the ratio of the volumes and scales the comparison to a per-parameter basis.

For the individual model optimal designs in all examples of this section save the last, Fedorov's algorithm via PROC OPTEX was run 50 times from randomly chosen initial designs and the best final design was chosen. For the final example, the MRMF algorithm was used to find the best designs for the models individually. Furthermore, the model-robust designs produced in this section, as well as those based upon DuMouchel and Jones (1994), were also generated based on 50 separate algorithm instances.

Example 1: Constrained Response Surface Experiment

A constrained two-factor example, taken from Heredia-Langner et al. (2004), will serve as an initial example illustrating our method. The design region, shown in Figure 1, is

$\chi = \{\mathbf{x} = (x_1, x_2) : -1 \leq x_1, x_2 \leq 1, x_1 + x_2 \leq 1, -0.5 \leq x_1 + x_2\}$, $n = 6$ and the experimenter would like a design robust for a first-order, a first-order with interaction, or full quadratic polynomial; i.e. $\mathcal{F} = \{\mathbf{f}_i^{'}(\mathbf{x}), 1 \leq i \leq 3, \mathbf{x} \in \chi\}$ where

$$\mathbf{f}_1^{'}(\mathbf{x}) = (1, x_1, x_2) \tag{7}$$

$$\mathbf{f}_2^{'}(\mathbf{x}) = (1, x_1, x_2, x_1x_2) \tag{8}$$

$$\mathbf{f}_3^{'}(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2) \tag{9}$$

The candidate list for this example consisted of 266 points constituting a grid of resolution 0.1 placed over the design space. We give two designs using the method of DuMouchel and Jones, Bayes 1 with $\mathbf{f}_{pri}^{'} = (1, x_1, x_2, x_1x_2)$ and $\mathbf{f}_{pot}^{'} = (x_1^2, x_2^2)$, and Bayes 2 with $\mathbf{f}_{pri}^{'} = (1, x_1, x_2)$ and $\mathbf{f}_{pot}^{'} = (x_1x_2, x_1^2, x_2^2)$. For both, we adopt $\frac{1}{\tau} = 1$. We also include in our comparison the model-robust design of Heredia-Langner et al. (2004) as well as the optimal design for the largest model.

The model-robust designs are shown in Figure 1. Two design points are common to all five designs, $(0, 1)$ and $(1, 0)$, and the MRMF and Bayes 1 designs are the same. Table 1 also compares the designs in terms of the determinant and $\mathcal{D}$-efficiency for each of the considered models, and the last column gives the product. The last row gives the determinant of the information matrix for the $\mathcal{D}$-optimal design for each of the models individually, and the efficiencies are calculated using these values.

Even though the Bayes 1 and MRMF designs seem close to the optimal design for the quadratic model (since their $\mathcal{D}$-efficiency for the quadratic model is nearly 1), the optimal-for-quadratic produces a poor design with respect to the interaction model. The Bayes 2 design, using only the linear terms as primary, produces a design with even worse performance for the interaction model. It is also somewhat surprising that the Bayes 1 design was the same as the MRMF design, given that we are considering three models. However, in this simple example the MRMF design for the three models is the same as that obtained when considering only models (8) and (9) and ignoring (7). Therefore, it

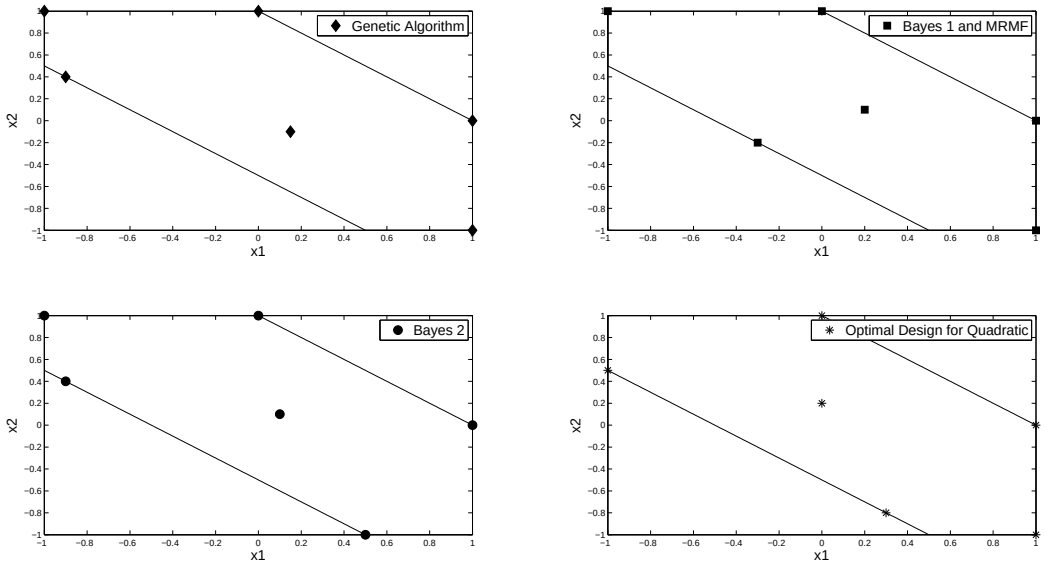

Figure 1: Model-robust Designs for Example 1

appears that the first-order model has no effect on the MRMF algorithm, so that there are essentially two models under consideration, a situation for which the Bayesian procedure is natural.

Example 2: Hypothetical Constrained 3-factor Experiment

To further explore our method and how it compares to the Bayesian method in particular, consider a three-factor example with design region $\chi = \{\mathbf{x} = (x_1, x_2, x_3) : -1 \leq x_1, x_2, x_3 \leq 1, -1 \leq x_1 + x_2 + x_3 \leq 1, -1 \leq x_1 + x_2 \leq 1, -1 \leq x_1 + x_3 \leq 1, -1 \leq x_2 + x_3 \leq 1\}$ and five

		Model			
Design	Measure	(7)	(8)	(9)	Product
MRMF	Determinant	27.04	33	3.01	2685.88
	$E_{\mathbf{f}}$	.810	.907	.995	.731
Genetic Algorithm	Determinant	31.14	26.91	2.21	1851.93
	$E_{\mathbf{f}}$	.849	.862	.945	.692
Bayes 1 ($\frac{1}{\tau} = 1$, primary terms those in (8))	Determinant	27.04	33	3.01	2685.88
	$E_{\mathbf{f}}$	.810	.907	.995	.731
Bayes 2 ($\frac{1}{\tau} = 1$, primary terms those in (7))	Determinant	32.33	13.15	3.08	1309.21
	$E_{\mathbf{f}}$	.860	.721	.998	.619
Optimal Design for (9)	Determinant	31.63	14.35	3.11	1411.60
	$E_{\mathbf{f}}$	.853	.737	1	.629
Optimal (for each model)	Determinant	50.88	48.77	3.11	

Table 1: Determinants, with $\mathcal{D}$-efficiencies, for Example 1 with $n = 6$, protecting against three models.

models of interest:

$$\mathbf{f}_1'(\mathbf{x}) = (1, x_1, x_2, x_3) \tag{10}$$

$$\mathbf{f}_2'(\mathbf{x}) = (\mathbf{f}_1', x_1x_2, x_1x_3, x_2x_3) \tag{11}$$

$$\mathbf{f}_3'(\mathbf{x}) = (\mathbf{f}_2', x_1^2, x_2^2, x_3^2) \tag{12}$$

$$\mathbf{f}_4'(\mathbf{x}) = (\mathbf{f}_3', x_1^2x_2, x_1^2x_3, x_1x_2^2, x_2^2x_3, x_1x_3^2, x_2x_3^2, x_1x_2x_3) \tag{13}$$

$$\mathbf{f}_5'(\mathbf{x}) = (\mathbf{f}_4', x_1^3, x_2^3, x_3^3) \tag{14}$$

so that $\mathcal{F} = \{\mathbf{f}_i'(\mathbf{x}), 1 \leq i \leq 5, \mathbf{x} \in \chi\}$.

In particular, assume that the experimenter would like to use $n = 20$ runs and would like a design that can fit each of these models well. To specify the Bayesian procedure, we take as primary all terms in (12) and designate the rest as potential. We give the MRMF design in Table 2, as well as Bayesian designs with $\frac{1}{\tau} = 1$ and $\frac{1}{\tau} = 16$ and the optimal design for the largest model, all using a candidate list consisting of a grid of points with resolution 0.1 placed over the design space.

The Bayesian designs are competitive, based on the $\mathcal{D}$-criterion, for most of the models but lack efficiency for model (13) when compared to the MRMF design. This might be

expected since it is in between the primary and full model and as such not explicitly considered. None of the designs perform very well for model (11), though the MRMF design is marginally better. As we expect, when a larger prior precision value is used in the Bayesian procedure, the efficiency of models containing primary terms is reduced, and in this case significantly degrades the design in terms of the product criterion. The optimal design for the largest model is competitive with the Bayesian designs in terms of model-robustness, though the MRMF design would likely be preferred because of its higher efficiencies in models (11), (12), and (13).

		Model					
Design	Measure	(10)	(11)	(12)	(13)	(14)	Product
MRMF	Determinant	6.58e3	5.57e4	1.10e5	3.21e0	5.24e-3	6.78e11
	$E_{\mathbf{f}}$	.864	.756	.870	.955	.979	.531
Bayes ($\frac{1}{\tau}=1$)	Determinant	6.63e3	5.21e4	9.74e4	9.92e-1	7.94e-3	2.65e11
	$E_{\mathbf{f}}$	.867	.749	.860	.892	.999	.498
Bayes ($\frac{1}{\tau}=16$)	Determinant	5.93e3	4.39e4	1.12e5	4.61e-1	4.41e-3	5.93e10
	$E_{\mathbf{f}}$	.843	.731	.872	.852	.970	.444
Optimal for (14)	Determinant	6.44e3	4.94e4	9.62e4	7.63e-1	8.07e-3	1.88e11
	$E_{\mathbf{f}}$	.860	.744	.859	.878	1	.483
Optimal (for each model)	Determinant	1.18e4	3.93e5	4.42e5	6.97e0	8.07e-3	

Table 2: Determinants, with $\mathcal{D}$-efficiencies, for Example 2 with $n = 20$, protecting against five models.

Example 3: Constrained Mixture Experiment

We now revisit the example (Heinsman and Montgomery 1995) briefly described at the outset. This is a four-factor constrained mixture experiment regarding the formulation of a household product in which 20 runs are available. The design region can be defined as:

$$\chi = \left\{ \mathbf{x} = (x_1, x_2, x_3, x_4) : \sum_{i=1}^{4} x_i = 1, 0.5 \leq x_1 \leq 1, 0 \leq x_2, x_3 \leq 0.5, 0 \leq x_4 \leq 0.05 \right\} \quad (15)$$

where x_1 is a nonionic surfactant, x_2 is an anionic surfactant, x_3 is a second nonionic surfactant, and x_4 is a zwitterionic surfactant. Because of the dependency induced by the mixture constraint, standard mixture design models are considered which do not include

an intercept:

$$\mathbf{f}_1^{'}(\mathbf{x}) = (\{x_i, i = 1, \ldots, 4\}) \tag{16}$$

$$\mathbf{f}_2^{'}(\mathbf{x}) = (\mathbf{f}_1^{'}, \{x_i x_j, i < j \leq 4\}) \tag{17}$$

$$\mathbf{f}_3^{'}(\mathbf{x}) = (\mathbf{f}_2^{'}, \{x_i x_j x_k, i < j < k \leq 4\}) \tag{18}$$

$$\mathbf{f}_4^{'}(\mathbf{x}) = (\mathbf{f}_3^{'}, \{x_i x_j (x_i - x_j), i < j \leq 4\}) \tag{19}$$

so that $\mathcal{F} = \{\mathbf{f}_i^{'}(\mathbf{x}), 1 \leq i \leq 4, \mathbf{x} \in \chi\}$. Heredia-Langner et al. (2004) also used this example, and so we compare our method to their Genetic Algorithm as well as to the Bayesian method of DuMouchel and Jones (1994). For the latter, we present designs using both a standard value for the prior precision ($\frac{1}{\tau} = 1$) and a larger precision value ($\frac{1}{\tau} = 16$), and primary terms those in (16) as well as (18). The potential terms depend upon the assignment of the primary terms and in each case are those unique to (19).

Since this is a large mixture design, we supplemented a regular grid (resolution 0.01) with extreme vertices and approximate centroids of the design region using code as described in Piepel (1988).

		Model				
Design	Measure	(16)	(17)	(18)	(19)	Product
MRMF	Determinant	5.31e-2	7.22e-22	2.65e-43	8.36e-78	8.49e-143
	$E_{\mathbf{f}}$	.728	.897	.931	.996	.606
Genetic Algorithm	Determinant	5.23e-2	7.46e-22	2.90e-43	7.80e-78	8.83e-143
	$E_{\mathbf{f}}$	.725	.900	.937	.992	.607
Bayes 1 ($\frac{1}{\tau} = 1$, primary terms those in (16)[1])	Determinant	5.46e-2	6.74e-22	2.24e-43	9.08e-78	7.48e-143
	$E_{\mathbf{f}}$	.733	.890	.919	1	.600
Bayes 2 ($\frac{1}{\tau} = 16$, primary terms those in (18))	Determinant	5.64e-2	6.12e-22	3.01e-43	3.08e-78	3.20e-143
	$E_{\mathbf{f}}$	.739	.882	.939	.947	.580
Bayes 3 ($\frac{1}{\tau} = 16$, primary terms those in (16))	Determinant	6.01e-2	6.07e-22	1.94e-43	3.87e-78	2.74e-143
	$E_{\mathbf{f}}$	.751	.881	.910	.958	.577
Optimal Design for (19)	Determinant	5.46e-2	6.74e-22	2.24e-43	9.08e-78	7.48e-143
	$E_{\mathbf{f}}$	.733	.890	.919	1	.600
Optimal (for each model)	Determinant	1.89e-1	2.15e-21	7.26e-43	9.08e-78	

Table 3: Determinant function values, with $\mathcal{D}$-efficiencies, for Example 3 with $n = 20$, protecting against four models.

[1]Same design results if primary terms are those in (18)

In Table 3, our method can be seen to be competitive with the Genetic Algorithm, though their design is slightly superior by our product optimality criterion. This is likely a function of the discretization in our candidate list. Note that the optimal design for model (19) has a significantly higher objective function value than that given in Heredia-Langner et al. (2004) (9.08e-78 vs. 7.83e-78), though theirs was asserted to have been obtained from PROC OPTEX in SAS as well.

It is the case again in this example that the best design found by the MRMF method is relatively close to that of the optimal design for the largest model. The Bayesian procedure, when the precision is 1, produces a design optimal for the largest model (Bayes 1 in Table 3) which is competitive using the product of the determinants as a criterion. The design is the same whether the primary terms are those in (16) or (18). When the precision is increased to 16, the Bayesian designs are different for different sets of primary terms (Bayes 2 and Bayes 3 in Table 3), but we see the same behavior as was noted before: The Bayesian designs become less efficient for the model that involves potential terms. The designs have slightly less variable $\mathcal{D}$-efficiencies, but suffer against the product optimality criterion.

Example 4: Mixture Experiment with Disparate Models

For our final example we use an unconstrained mixture experiment by Frisbee and McGinity (1994) with $n = 11$. The response is the glass transition temperature of a certain film with three nonionic surfactant factors. The goal was to minimize this transition temperature, and Frisbee and McGinity fit a traditional polynomial model. However, another class of models, the so-called Becker models (Cornell 1990, Sec. 6.5), were shown by Rajagopal and del Castillo (2005) to also fit the data well and lead to a significantly different optimal solution. These models, originally considered to address certain shortcomings in the Sheffé polynomial models, model second-order mixture blending of factors using $\min(\cdot)$. For instance, for factors i and j, $\min(x_i, x_j)$ is used instead of $x_i x_j$.

In this case,

$$\chi = \left\{ \mathbf{x} = (x_1, x_2, x_3) : \sum_{i=1}^{3} x_i = 1, 0 \leq x_i \leq 1, i = 1, 2, 3 \right\} \tag{20}$$

and we take five possible models:

$$\mathbf{f}_1^{'}(\mathbf{x}) = (\{x_i, i = 1, 2, 3\}) \tag{21}$$

$$\mathbf{f}_2^{'}(\mathbf{x}) = (\mathbf{f}_1^{'}, \{x_i x_j, i < j \leq 3\}) \tag{22}$$

$$\mathbf{f}_3^{'}(\mathbf{x}) = (\mathbf{f}_2^{'}, \{x_1 x_2 x_3\}) \tag{23}$$

$$\mathbf{f}_4^{'}(\mathbf{x}) = (\mathbf{f}_1^{'}, \{min(x_i, x_j), i < j \leq 3\}) \tag{24}$$

$$\mathbf{f}_5^{'}(\mathbf{x}) = (\mathbf{f}_4^{'}, \{min(x_1, x_2, x_3)\}) \tag{25}$$

so that $\mathcal{F} = \{\mathbf{f}_i^{'}(\mathbf{x}), 1 \leq i \leq 5, \mathbf{x} \in \chi\}$.

In addition to the five models we are guarding against, we also examine effectiveness of our design with respect to the model fit by Frisbee and McGinity, as well as the most probable model found *a posteriori* by Rajagopal and Castillo:

$$\mathbf{f}_{fm}^{'}(\mathbf{x}) = (x_1, x_2, x_3, x_1 x_3, x_2 x_3) \tag{26}$$

$$\mathbf{f}_{rc}^{'}(\mathbf{x}) = (x_1, x_2, x_3, min(x_1, x_3), min(x_2, x_3)) \tag{27}$$

For a candidate list, we used a regular grid with resolution 1/12, which because of the regular design region, contained the vertices and centroids of the region.

With the disparate model types, the Bayes procedure, with its primary and potential factors, cannot be easily applied. Instead, we examine the results of the MRMF design and compare it in Table 4 to the design that was actually used. In terms of efficiency, the design used by Frisbee and McGinity is much inferior for all models considered because it includes, in addition to two centroid points, three other points on the interior of the simplex design region.

As seen in Table 4, the MRMF design is optimal for models (23), (24), and (25). This is because the optimal designs for these models individually are interchangeable; i.e. the optimal design for one is also optimal for another. Note that since the models are not nested we do not have the multiresponse $\mathcal{D}$-optimality interpretation.

		Model						
Design	Measure	(21)	(22)	(23)	(24)	(25)	(26)	(27)
MRMF	Determinant	19.81	5.91e-3	5.36e-6	0.569	2.78e-2	6.61e-2	1.46
	$E_{\mathbf{f}}$	.745	.954	1	1	1	.812	.866
Frisbee and McGinity	Determinant	8.25	1.22e-3	1.51e-6	.146	8.82e-3	2.30e-2	.588
	$E_{\mathbf{f}}$	.556	.733	.835	.797	.849	.658	.722
Optimal (for each model)	Determinant	48	7.8e-3	5.36e-6	.569	2.73e-2	.188	3

Table 4: Objective Function Values and $\mathcal{D}$-efficiencies for Example 4, $n = 11$, protecting against 5 Models.

Model Spaces

The model spaces considered so far contain maximal models within certain classes. For instance, in Example 1, the model set we utilized included a model with all main effects (i.e. maximal main effects model), a model with all main effects and all two-factor interactions (i.e. maximal two-factor interaction model), and a model with all terms up through quadratic (i.e. maximal second-order polynomial model). In many standard experimental situations, a certain class of model is envisioned and designed for (i.e. a fractional factorial design when an interaction model is assumed), and in such cases a departure from this class might be hedged against by implementing a model-robust design as described in this paper.

However, the experimenter might wish to account for a much larger set of possible models. For instance, Li and Nachtsheim (2000) consider models composed of all main effects and up to g two-way interactions. In our setting we have assumed that higher-order terms are of interest as well, and this suggests an enormous model space even for moderately sized problems, if all possible models are contemplated. As the size of the problem increases, the space of all possible polynomial models grows exponentially in the number of *terms* in the largest model considered. As such, an explicit accounting for each one is infeasible, at least

using the present methodology. On the other hand, many of these models do not obey effect hierarchy (defined below), and limiting the possible models to those which do will make the model space much smaller, though still unmanageable for large problems.

The model sets used in this paper may be viewed as an approximation to these larger model spaces. But how good is this approximation? One might intuit that unless there is a high degree of certainty about the model, even our crude approximation is better than using a single model to specify the $\mathcal{D}$-optimal design. But how much better? And how much are we giving up by not explicitly considering all possible models?

We will consider these questions for two examples, necessarily small because of the computational requirements. This is an exploratory examination, not a comprehensive, definitive study and we note that other model space approximations might be chosen. For instance, Bingham and Chipman (2007) select sets of *a priori* model forms based upon a prior distribution of possible models using three axiomatic principles: Effect sparsity, effect hierarchy, and effect heredity. Their approach does not eliminate the problem of exploding model size (in the end, they only handle a small subset of all possible models—those with the largest prior probabilities), but it does represent a principled approach to choosing a set of possible models when the space of all possible models is too large. However, it would not be a straightforward adaptation to our setting because, unlike them, we consider quadratic and cubic effects.

In this section we consider two different model spaces: 1) All possible polynomial models up to quadratic (excluding the intercept-only model); and 2) All possible hierarchical polynomial models up to quadratic (also excluding the intercept-only model). The first has many models that would probably not be fit in practice (for instance, in a two-factor experiment, the model $\mathbf{f}^{'}(\mathbf{x}) = (1, x_1x_2, x_1^2)$), and so the hierarchical alternative may represent a more realistic set of possible models. We will evaluate the effectiveness of our approximation for each.

First, we define a hierarchical effect structure which includes quadratic terms.

Definition 1 *A* hierarchical effect structure *includes, in ascending order: main effects,*

two-factor interactions, and quadratic terms. Before two-factor interactions can appear, the corresponding main effects must be included. Before a quadratic effect can be considered, the corresponding main effect as well as at least one two-way interaction involving the appropriate factor must be present.

Thus, $\mathbf{f}'(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2)$ would be hierarchical, but $\mathbf{f}'(\mathbf{x}) = (1, x_1, x_2, x_1^2)$ would not. Higher order terms (for instance, three-way interactions or cubic terms) could also be included but are not in these examples.

Approximate Model Spaces for Example 1

To study the effectiveness of the sort of model space approximations that are used in this paper, we will compare the model-robust design for Example 1 found in the previous section with the model-robust designs using all possible models, the model-robust design using all possible hierarchical models, and the design optimal for the full quadratic model alone. All designs are based on 50 algorithm tries and the candidate list described in the previous section. There are seven hierarchical models and $2^5 - 1 = 31$ possible models (we exclude the intercept-only model, in both cases).

We first evaluate each of the designs in terms of the average $\mathcal{D}$-efficiencies with respect to all 31 possible models (Table 5), also noting the standard deviations and minimum of the $\mathcal{D}$-efficiencies. The model-robust design using just three models (3-MR in the table) loses about 2.2% on average when compared to the design that is robust for all 31 models, with a slightly smaller standard deviation and minimum. It is, however, much better than the design optimal for the quadratic model alone by each measure.

Design	Mean	StDev	Min
3-MR	0.875	0.071	0.697
31-MR	0.896	0.073	0.708
Quadratic Only	0.813	0.102	0.567

Table 5: Comparison of model-robust designs for Example 1, in terms of $\mathcal{D}$-efficiencies with respect to all 31 possible models.

When evaluated with respect to the seven hierarchical models, Table 6 shows that the

3-MR design now has an average $\mathcal{D}$-efficiency 0.045 less than the design robust to all 7 hierarchical models, while still 0.035 better than the quadratic only design.

Design	Mean	StDev	Min
3-MR	0.836	0.107	0.697
7-MR	0.881	0.106	0.731
Quadratic Only	0.801	0.109	0.684

Table 6: Comparison of model-robust designs for Example 1, in terms of $\mathcal{D}$-efficiencies with respect to all 7 possible hierarchical models.

Approximate Model Spaces for a 3-factor Experiment

The second example is a generic unconstrained 3-factor experiment for which $\chi = \{\mathbf{x} = (x_1, x_2, x_3) : -1 \leq x_1, x_2, x_3 \leq 1\}$ and again models as large as full quadratic are considered. Since for the full second-order model there are 9 non-intercept terms, there are $2^9 - 1 = 511$ possible models. In addition, using Definition 1, there are 63 possible hierarchical models.

In Table 7, we evaluate several designs using the mean, standard deviation, and minimum $\mathcal{D}$-efficiency with respect to the 511 possible models. In Table 8, we do the same with respect to the 63 hierarchical models. To calculate $\mathcal{D}$-efficiencies, the individually optimal designs for each of the 511 possible models were found, each using 50 algorithm tries and a 3^3 candidate list.

We found three different model-robust designs, one using just three models (3-MR, the model set being the main effects only model, main effects plus 2-factor interaction model, and full quadratic), another robust with respect to the 63 hierarchical models, and another robust with respect to all 511 possible models. We found that the model-robust designs potentially required a denser grid of candidate points, and consequently used one with a resolution of 0.1 (21^3 = 9,261 total points). This, along with the large size of the model set significantly slowed computations and required us to make evaluations of these three designs based upon only 2 algorithm tries.

Table 7 shows that when measured against all 511 possible models, the 3-MR design as well as the design robust for all hierarchical models are slightly inferior to both the 511-MR

Design	Mean	StDev	Min
3-MR	0.864	0.060	0.576
63-MR	0.865	0.060	0.579
511-MR	0.877	0.069	0.705
Quadratic Only	0.872	0.069	0.690

Table 7: Comparison of model-robust designs for 3-factor experiment, in terms of $\mathcal{D}$-efficiencies with respect to all 511 possible models. Candidate list for model-robust designs was the 21^3 set of points.

and quadratic only designs, in terms of average $\mathcal{D}$-efficiency. The contrast between the designs is more stark when the minimum $\mathcal{D}$-efficiency is considered, however. The 3-MR and 63-MR designs often suffer for nonhierarchical models for which quadratic terms are included with no, or few, lower-order terms. For instance, the 3-MR design is only 57.6% $\mathcal{D}$-efficient with respect to the model $\mathbf{f}'(\mathbf{x}) = (1, x_1^2, x_2^2, x_3^3)$. If the seven models which include just an intercept and quadratic terms (with no lower order effects) are eliminated, the minimum $\mathcal{D}$-efficiency becomes 0.67.

Design	Mean	StDev	Min
3-MR	0.907	0.033	0.835
63-MR	0.907	0.034	0.833
Quadratic Only	0.881	0.068	0.740

Table 8: Comparison of model-robust designs for 3-factor experiment, in terms of $\mathcal{D}$-efficiencies with respect to all 63 possible models. Candidate list for model-robust designs was the 21^3 set of points.

On the other hand, if only hierarchical models are considered, the tables are turned and the 3-MR and 63-MR designs have a higher average $\mathcal{D}$-efficiency, by 0.026, with much smaller standard deviations and nearly 0.1 advantage on the minimum. Since the 3-MR design is based upon hierarchical models, this outcome is not surprising, though belied somewhat by the previous example.

We stress that these are preliminary, noncomprehensive results and general conclusions should not be drawn from them. However, they do provide some evidence that, at least in the hierarchical case, a crude approximation of the model space is superior to the quadratic-only optimal design.

Discussion

The Model-robust Modified Fedorov (MRMF) exchange algorithm presented in this paper provides a natural tool with which to find designs when an optimal design is desired but the model-form is unknown. The mechanism to achieve this is intuitive and simple: The experimenter chooses r models for which he/she would like to design. Then, a design is found which maximizes the product of the determinant of the information matrices of each of the models.

Furthermore, the MRMF method produces designs that are competitive, with simpler algorithmic machinery, than the Genetic Algorithm approach of Heredia-Langner et al. (2004). The strength of the MRMF method with respect to the GA technique is that it is automatic and a straightforward extension of commonly used exchange algorithms. The GA requires tuning of several parameters and is nontrivial to implement effectively.

We also compared our procedure to the Bayesian method of DuMouchel and Jones (1994), a widely available model-robust technique. We initially hypothesized that the Bayesian method would suffer when confronted with multiple possible models, since it categorizes terms into just two groups. This is supported by Example 2, though the procedure performed fairly well in Examples 1 and 3. The choice of $\frac{1}{\tau}$ certainly affects the model-robustness of the design; indeed for certain values of $\frac{1}{\tau}$ (i.e. $\frac{1}{\tau} = 1$ in Example 3) the method seems to produce a design optimal for the highest-order model, while for large enough values of $\frac{1}{\tau}$ the full model is not even estimable. The choice of terms as primary or potential also makes an impact. Our procedure does not suffer from these uncertainties, has a multiresponse $\mathcal{D}$-optimal interpretation (for $\mathcal{F}$ nested) and explicitly considers a larger class of models; it can also handle situations as in Example 4 in which the possible models are disparate and impossible to nest.

One strategy, if faced with a situation necessitating a $\mathcal{D}$-optimal design, might be to design for the highest-order model possible. If, as assumed in this paper, there are a sufficient number of runs to estimate the largest model, one might question whether the efficiency

gained in model-robust methods is worth the additional methodology. In certain cases, as in the third example, the gains appear to be limited. But as demonstrated by the first and second examples, significant gains can be made by utilizing the model-robust approach. Therefore, a dedicated procedure based upon accepted univariate exchange algorithms will be useful to produce model-robust designs.

In terms of $\mathcal{D}$-efficiency, the MRMF designs can be seen to favor larger models. In other words, the efficiencies of the smaller models suffer as compared to the larger ones. To mitigate this, one might consider the following optimization criterion (Atkinson et al. 2007, Emmett et al. 2007), instead of (5):

$$\boldsymbol{\phi}(\mathbf{M}_{\mathcal{F}}(\tilde{\xi}_n)) = \prod_{i=1}^{r} \left|\mathbf{M}_{\mathbf{f}_i}(\tilde{\xi}_n)\right|^{1/q_i} \tag{28}$$

where q_i is the number of parameters in the i^{th} model. It is straightforward to derive an exchange algorithm using this criterion–call it the scaled MRMF–which has the effect of shrinking values of dissimilar orders of magnitude toward each other, in essence weighting more heavily those models with fewer parameters. We implemented this procedure using several examples, and the results were surprisingly similar. For instance, for the constrained mixture experiment in Example 3, the scaled MRMF design resulted in a design very close to the MRMF in Table 3. For the hypothetical experiment in Example 2, we observed more of a difference, with $\mathcal{D}$-efficiencies increasing from 86.4% to about 89% for model (10) and from 75.6% to about 78% for model (11), while decreasing the efficiencies of model (12) from 87% to about 85.5% and model (14) from 97.9% to about 96.5%, but still resulting in a design with an imbalance in the $\mathcal{D}$-efficiencies.

The model-robust criterion used in this paper could easily be extended to include prior information in terms of model weights, if certain models are preferred over the others. However, since this work was motivated in part by multiresponse optimal design theory, the minimal volume of the parameter confidence ellipsoid interpretation of $\mathcal{D}$-optimality is used and thus we only consider equally weighted models. Furthermore, the relative ineffectiveness

of the scaled MRMF to provide designs with similar $\mathcal{D}$-efficiencies for all considered models underscores the difficulty in balancing the $\mathcal{D}$-efficiencies using weights.

We also addressed the issue alluded to by a referee, that the model sets used in our examples are small compared to the set of all possible models. We examined two small examples and found that though general conclusions cannot be drawn, for the all possible hierarchical models at least, the small sets of the type used in this paper offer improvements over single-model $\mathcal{D}$-optimal designs and are competitive compared to the designs robust for all possible hierarchical models in certain situations.

Finally, assume that T_e is the time it takes to run the univariate Modified Fedorov exchange algorithm. The runtime for these model-robust algorithms should be rT_e where r is the number of models considered. Commercial software programs have fast implementations of this exchange algorithm, so the computational burden imposed by a similarly implemented model-robust exchange algorithm, for many problems, should not be heavy. However, a candidate set-free version of our model-robust algorithm would be a significant computational improvement. The coordinate exchange algorithm (Meyer and Nachtsheim 1995) does not require a candidate set, but fails to easily accommodate mixture constraints. There is ongoing work to develop an effective candidate set-free algorithm to handle such constraints (an attempt is given by Piepel et al. 2005), but a satisfactory procedure has yet to appear in the literature.

Note: All designs referred to in the Examples section, as well as Matlab code to generate the MRMF designs, are available at `http://www.stat.psu.edu/~jlr/pub/Smucker-JQT/`. In addition, the model-robust designs alluded to in the Model Spaces section are included as well.

Acknowledgements

The authors would like to thank Greg Piepel for furnishing us with a usable electronic version of his Fortran code for finding extreme vertices and approximate centroids for constrained

mixture design regions, as well as his comments on an earlier version of this paper. We are also grateful to two anonymous referees whose reviewed provoked improvements to this manuscript.

References

Atashgah, A. B. and Seifi, A. (2009). Optimal design of multi-response experiments using semi-definite programming. *Optimization and Engineering*, 10(1):75–90.

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press.

Bingham, D. R. and Chipman, H. A. (2007). Incorporating prior information in optimal design for model selection. *Technometrics*, 49(2):155–163.

Bischoff, W. (1993). On d-optimal designs for linear models under correlated observations with an application to a linear model with multiple response. *Journal of Statistical Planning and Inference*, 37:69–80.

Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of a response surface design. *Journal of the American Statistical Association*, 54(287):622–654.

Chang, S. I. (1997). An algorithm to generate near d-optimal designs for multiple-response surface models. *IIE Transactions*, 29:1073–1081.

Chang, Y.-J. and Notz, W. I. (1996). *Handbook of Statistics, Vol. 13*, chapter Model Robust Designs, pages 1055–1098. Elsevier Science B. V.

Cook, R. D. and Nachtsheim, C. J. (1980). A comparison of algorithms for constructing exact d-optimal designs. *Technometrics*, 22:315–324.

Cook, R. D. and Nachtsheim, C. J. (1982). Model-robust, linear-optimal designs. *Technometrics*, 24:49–54.

Cornell, J. (1990). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*. Wiley-Interscience, 2nd edition.

Dette, H. (1990). A generalization of $\mathcal{D}$- and $\mathcal{D}_1$-optimal design in polynomial regression. *Annals of Statistics*, 18:1784–1804.

Dette, H. and Franke, T. (2001). Robust designs for polynomial regression by maximizing a minimum of $\mathcal{D}$- and $\mathcal{D}_1$-efficiencies. *The Annals of Statistics*, 29(4):1024–1049.

DuMouchel, W. and Jones, B. (1994). A simple bayesian modification of d-optimal designs to reduce dependence on an assumed model. *Technometrics*, 36(1):37–47.

Emmett, M., Goos, P., and Stillman, E. (2007). D-optimal designs for multiple polynomial responses. Technical report, University of Sheffield.

Fang, Z. and Wiens, D. P. (2000). Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximate linear models. *Journal of the American Statistical Association*, 95(451):807–818.

Fang, Z. and Wiens, D. P. (2003). Robust regression designs for approximate polynomial models. *Journal of Statistical Planning and Inference*, 117:305–321.

Fedorov, V. V. (1972). *Theory of Optimal Experiments.* Academic Press, New York, NY.

Frisbee, S. E. and McGinity, J. W. (1994). Influence of nonionic surfactants on the physical and chemical properties of a biodegradable pseudolatex. *European Journal of Pharmaceutics and Biopharmaceutics*, 40:355–363.

Goos, P., Kobilinsky, A., O'Brien, T. E., and Vandebroek, M. (2005). Model-robust and model-sensitive designs. *Computational Statistics & Data Analysis*, 49:201–216.

Heinsman, J. A. and Montgomery, D. C. (1995). Optimization of a household product formulation using a mixture experiment. *Quality Engineering*, 7:583–600.

Heredia-Langner, A., Montgomery, D. C., Carlyle, W. M., and Borror, C. M. (2004). Model-robust optimal designs: A genetic algorithm approach. *Journal of Quality Technology*, 36(3):263–279.

Johnson, M. E. and Nachtsheim, C. J. (1983). Some guidelines for constructing exact d-optimal designs on convex design spaces. *Technometrics*, 25:271–277.

Jones, B., Lin, D. K. J., and Nachtsheim, C. J. (2008). Bayesian d-optimal supersaturated designs. *Journal of Statistical Planning and Inference*, 138:86–92.

Jones, B. A., Li, W., Nachtsheim, C. J., and Ye, K. Q. (2009). Model-robust supersaturated and partially supersaturated designs. *Journal of Statistical Planning and Inference*, 139:45–53.

Khuri, A. I. and Cornell, J. A. (1996). *Response Surfaces: Designs and Analyses.* M. Dekker.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society, Series B*, 21:272–304.

Kurotschka, V. G. and Schwabe, R. (1996). *Madan Puri Festschrift*, chapter The Reduction of Design Problems for Multivariate Experiments to Univariate Possibilities and their Limitations, pages 193–204. VSP.

Läuter, E. (1974). Experimental design in a class of models. *Statistics*, 5:379–398.

Li, W. and Nachtsheim, C. J. (2000). Model-robust factorial designs. *Technometrics*, 42(4):345–352.

Meyer, R. K. and Nachtsheim, C. J. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37(1):60–69.

Mitchell, T. J. (1974). An algorithm for the construction of d-optimal experimental designs. *Technometrics*, 16:203–210.

Montepiedra, G. and Fedorov, V. V. (1997). Minimum bias designs with constraints. *Journal of Statistical Planning and Inference*, 63:97–111.

Neff, A. (1996). *Bayesian two-stage designs under model uncertainty.* PhD thesis, Virginia Polytechnic Institute and State University.

Piepel, G. F. (1988). Programs for generating extreme vertices and centroids of linearly constrained experimental regions. *Journal of Quality Technology*, 20:2:125–139.

Piepel, G. F., Cooley, S. K., and Jones, B. (2005). Construction of a 21-component layered mixture experiment design using a new mixture coordinate-exchange algorithm. *Quality Engineering*, 17:579–594.

Rajagopal, R. and del Castillo, E. (2005). Model-robust process optimization using bayesian model averaging. *Technometrics*, 47(2):152–163.

SAS (2000-2004). SAS Institute Inc., Version 9.1.3, Cary, NC, USA.

Sun, D. X. (1993). *Estimation capacity and related topics in experimental designs.* PhD thesis, University of Waterloo.

Welch, W. J. (1983). A mean squared error criterion for the design of experiments. *Biometrika*, 70(1):205–213.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57:348–368.

Zhou, J. (2008). D-optimal minimax regression designs on discrete design space. *Journal of Statistical Planning and Inference*, 138:4081–4092.