Statistica Sinica **15**(2005), 619-633

DESIGNING EXPERIMENTS UNDER RANDOM CONTAMINATION WITH APPLICATION TO POLYNOMIAL SPLINE REGRESSION

D. C. Woods

University of Southampton

Abstract: In science and engineering, there is often uncertainty in the linear model assumed for a response when an experiment is being designed. The errors in predictions made from a fitted model may then be more dependent on the systematic errors (bias) that arise from model misspecification than from errors related to the variance of the estimators of the unknown model coefficients. This may result in commonly used criteria, such as $D-$optimality, being inappropriate for the selection of a design. In this paper the true model is allowed to differ from the model assumed for design purposes by an additive contamination term which is a random variable. Design criteria are defined which involve properties of the resulting random bias. These criteria are applied to assumed linear models constructed from polynomials or from polynomial splines where the contamination is modeled by random variables that represent the unknown numbers, locations and coefficients of additional knots (or breakpoints). Designs are found using an exchange algorithm which incorporates Monte Carlo simulation to approximate properties of the bias distribution. When the expectation of the contamination is zero, theoretical results enable a reduction in the computationally intensive design search. The extension of the approach to the use of a mean squared error criterion is also considered.

Key words and phrases: Bias, design selection criteria, exchange algorithm, mean squared error, Monte Carlo simulation, optimal design, polynomial splines.

1. Introduction

Suppose that, prior to experimentation, the relationship between a continuous response Y and m factors or variables $\boldsymbol{x} = (x_1, \ldots, x_m)$ is believed to be described, over the design region, by an *assumed model*

$$Y(\boldsymbol{x}) = f(\boldsymbol{x}) + \varepsilon\,, \tag{1.1}$$

where $f(\cdot)$ is a function of the factors and ε is a random error variable. The random errors are assumed to be independently and identically distributed with zero mean and variance σ^2. Suppose also that the experimenter's prior beliefs about the form of f are incorrect, and that the *true model* is

$$Y(\boldsymbol{x}) = f(\boldsymbol{x}) + \phi(\boldsymbol{x}) + \varepsilon\,, \tag{1.2}$$

where $\phi(\boldsymbol{x})$ represents the misspecification of the assumed model and is called the *contamination*.

In practical experiments, the form of $\phi(\boldsymbol{x})$ is often unknown. The contamination might be produced by one or more unidentified factors that arise, for example, from a production process or from unforeseen calibration errors in laboratory measurements. If the presence of these factors had been anticipated when the experiment was being planned, then their effects could have been included in the assumed model and, where control of the factor levels is possible, in the design of the experiment. Also, there may be some additional factors or model terms, such as quadratic effects, which the experimenter might have considered but (wrongly) decided not to address in planning the experiment.

A further situation, which often occurs in practical applications in the physical sciences, is where simple models that approximate theoretical relationships are fitted to experimental data in order to obtain empirical estimates of quantities of scientific interest. By their very nature these approximations are often subject to misspecification and so the models are fitted to the data in the presence of, possibly unknown, contamination.

If the contamination $\phi(\boldsymbol{x})$ were known, then a design could be found using the criterion of Box and Draper (1959) which minimizes the integrated mean square error (IMSE) over the design region. The IMSE is the sum of the average normalized prediction *variance* (V), representing random errors, and the average normalized squared prediction *bias* (B), representing systematic errors due to model inadequacy. That is,

$$\begin{aligned} IMSE &= \frac{n\Omega}{\sigma^2}\int_{\mathcal{R}} E\left\{[\hat{f}(\boldsymbol{x}) - E(\hat{f}(\boldsymbol{x}))]^2\right\}\,\mathrm{d}\boldsymbol{x} \\ &\qquad + \frac{n\Omega}{\sigma^2}\int_{\mathcal{R}} \left\{E[\hat{f}(\boldsymbol{x})] - E[Y(\boldsymbol{x})])\right\}^2\,\mathrm{d}\boldsymbol{x} \\ &= V + B(\phi)\,, \end{aligned} \tag{1.3}$$

where n is the number of points in the design, $\Omega = (\int_{\mathcal{R}} \mathrm{d}x)^{-1}$, $\mathcal{R}$ is the design region of interest and $\hat{f}$ is the estimator of f obtained by fitting (1.1). Model (1.2) is most appropriate when, prior to experimentation, V and $B(\phi)$ are considered to be of similar magnitude. If $B(\phi)$ is thought to be much larger than V, then an alternative to f should be sought. If V is much larger than $B(\phi)$, then f is a good approximation to the true function and the planning of the experiment can proceed under the assumption that (1.1) is correct.

Most work in the literature on minimum bias designs concerns polynomial regression models with $\phi(\boldsymbol{x})$ consisting of a linear combination of known higher order monomials and interactions, that is $\phi(\boldsymbol{x}) = \boldsymbol{x}_c'\boldsymbol{\gamma}$, where $\boldsymbol{x}_c$ is a vector of known functions of $\boldsymbol{x}$ and $\boldsymbol{\gamma}$ is a vector of unknown coefficients. In the pioneering

work of Box and Draper (1959) an example was presented where, even when V was four times as large as B, the design that minimized the IMSE was closer to the *all-bias* design that minimizes B than to the V-optimal design that minimizes V alone.

Notz (1989) and Allen, Yu and Schmitz (2003) modeled $\boldsymbol{\gamma}$ as a random vector with known covariance structure. Notz (1989) assumed that the elements of $\boldsymbol{\gamma}$ were independent with zero mean and known variance and used convex design theory (see, for example, Fedorov (1972)) to find designs for a single factor that minimized the expectation of the determinant or trace of a mean squared error matrix. Allen, Yu and Schmitz (2003) assumed a known covariance structure for $\boldsymbol{\gamma}$ and used an expectation criterion similar to that described in Section 6 of this paper to find designs for a die casting process using a genetic algorithm. Alternative approaches to finding designs in the presence of model contamination include those of Montepiedra and Fedorov (1997), Welch (1983), Wiens and Zhou (1997) and Heo, Schmuland and Wiens (2001). The work in this paper adds an additional level of model uncertainty to that in the literature by assuming that both $\boldsymbol{x}_c$ and $\boldsymbol{\gamma}$ are unknown.

We represent our uncertainty about the model misspecification by modeling $\phi(\boldsymbol{x})$ in (1.2) as a realization of a random variable, $\Phi(\boldsymbol{x})$, with a known distribution. This distribution may be chosen by using any available prior information on the likely nature and size of the contamination. Thus a *population* of true models is considered, given by

$$Y(\boldsymbol{x}) = f(\boldsymbol{x}) + \Phi(\boldsymbol{x}) + \varepsilon, \tag{1.4}$$

where $\Phi(\boldsymbol{x})$ and ε are independent random variables.

The aim of this paper is to find designs that offer protection against bias arising from models (1.4). Design selection criteria based on properties of the distribution of the bias are investigated and implemented using a search algorithm which incorporates Monte Carlo simulation. The methodology presented applies to any general linear model.

In Section 2, the bias for a model with random contamination is defined and an expression is obtained for random bias under a linear model. Section 3 defines three design criteria and describes a design search algorithm for their implementation. In Section 4 the ideas are applied to a class of flexible linear regression models built from polynomial spline basis functions (see, for example, Eubank (1999, Chap.6)).

For the case when the expected contamination is zero, Section 5 gives a theorem which allows the computational efficiency of the design search to be

improved. In Section 6 the design selection criteria are extended to include the prediction variance, V, and some conclusions are given in Section 7.

2. Random Contamination and Bias

Suppose that a linear model $f(\boldsymbol{x}) = \boldsymbol{x}_a'\boldsymbol{\beta}$ is assumed, where $\boldsymbol{x}_a$ is a $q\times 1$ vector holding functions of $\boldsymbol{x}$, and $\boldsymbol{\beta}$ is the $q\times 1$ vector of unknown coefficients. The bias resulting from any given design and assumed model is a random variable, $B(\Phi)$, which takes the value $B(\phi)$ for a realization $\Phi(\boldsymbol{x}) = \phi(\boldsymbol{x})$, where, from (1.3),

$$B(\phi) = \frac{n\Omega}{\sigma^2}\int_{\mathcal{R}} \left[E(\boldsymbol{x}_a'\hat{\boldsymbol{\beta}}|\phi) - \boldsymbol{x}_a'\boldsymbol{\beta} - \phi(\boldsymbol{x})\right]^2 \,\mathrm{d}\boldsymbol{x}\,. \tag{2.1}$$

Let X be the $n\times q$ model matrix and $\boldsymbol{Y}$ be the $n\times 1$ vector of observations from the experiment. Then

$$E[\boldsymbol{x}_a'\hat{\boldsymbol{\beta}}|\phi] = \boldsymbol{x}_a'(X'X)^{-1}X'E(Y(\boldsymbol{x})) = \boldsymbol{x}_a'\boldsymbol{\beta} + \boldsymbol{x}_a'A\boldsymbol{\phi}_D\,, \tag{2.2}$$

where $A = (X'X)^{-1}X'$ and $\boldsymbol{\phi}_D$ is an $n\times 1$ vector of values of the contamination for each design point. From (2.1) and (2.2), each realization of the bias can be written as

$$B(\phi) = \frac{n\Omega}{\sigma^2}\int_{\mathcal{R}} \left[\boldsymbol{x}_a'A\boldsymbol{\phi}_D - \phi(\boldsymbol{x})\right]^2 \,\mathrm{d}\boldsymbol{x}\,. \tag{2.3}$$

Throughout this paper, the continuous design region $\mathcal{R}$ is approximated by an evaluation grid of r points containing the n design points. Equation (2.3) is then approximated by

$$B(\phi) = \frac{n}{r\sigma^2}\sum_{i=1}^{r}\left[\boldsymbol{x}_{a,i}'A\boldsymbol{\phi}_D - \phi(\boldsymbol{x}_i)\right]^2 = \frac{n}{r\sigma^2}\left[FAD\boldsymbol{\phi} - \boldsymbol{\phi}\right]'\left[FAD\boldsymbol{\phi} - \boldsymbol{\phi}\right]\,,$$

where F is an $r\times q$ matrix of assumed model terms for each evaluation point, $\boldsymbol{\phi}$ is an $r\times 1$ vector of contaminations for each evaluation point, and D is an $n\times r$ matrix with (i,j)th entry non-zero and equal to 1 only when the ith design point and the jth evaluation point coincide. The vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_{a,i}$ hold the m factor values and the q assumed model terms, respectively, for the ith evaluation point; see also Welch (1983) for non-random contamination. It follows that

$$B(\phi) = \frac{n}{r\sigma^2}\mathrm{tr}\left\{[FAD - I]'[FAD - I]\boldsymbol{\phi}\boldsymbol{\phi}'\right\}\,, \tag{2.4}$$

where I is an $r\times r$ identity matrix. Every realization of the random variable $B(\Phi)$ has the form (2.4) and hence the random bias is

$$B(\Phi) = \frac{n}{r\sigma^2}\mathrm{tr}\left\{[FAD - I]'[FAD - I]\boldsymbol{\Phi}\boldsymbol{\Phi}'\right\}\,,$$

where the jth entry of the $r \times 1$ random vector $\mathbf{\Phi}$ is the contamination random variable $\Phi(\boldsymbol{x}_j)$ $(j = 1, \ldots, r)$.

3. Bias-Based Criteria and Their Implementation

Let $\mathcal{D}$ denote the set of all n-point designs in $\mathcal{R}$, and $A(\delta)$ and $D(\delta)$ denote the respective matrices A and D evaluated for design $\delta \in \mathcal{D}$. Three types of optimal design are considered as follows.

- A design $\delta^* \in \mathcal{D}$ is *expected bias* $(EB-)$ *optimal* over $\mathcal{D}$ if $E\left[B(\Phi)|\delta^*\right] = \min_{\delta \in \mathcal{D}} E\left[B(\Phi)|\delta\right]$, where

$$E\left[B(\Phi)|\delta\right] = \frac{n}{r\sigma^2}\mathrm{tr}\left\{[FA(\delta)D(\delta) - I]'[FA(\delta)D(\delta) - I]E\left[\mathbf{\Phi}\mathbf{\Phi}'\right]\right\} . \quad (3.1)$$

- For given $0 < p \leq 1$, a design $\delta^* \in \mathcal{D}$ is said to be pth *percentile bias* $(PB(p)$-$)$ *optimal* over the set $\mathcal{D}$ if $b(\delta^*, p) = \min_{\delta \in \mathcal{D}} b(\delta, p)$, where $\Pr(B(\Phi) < b(\delta, p)|\delta) = p$. *Median bias optimality* is obtained for $p = 0.5$.
- A design $\delta^* \in \mathcal{D}$ is *variance bias* $(VB$-$)$ *optimal* over $\mathcal{D}$ if $\mathrm{Var}\left[B(\Phi)|\delta^*\right] = \min_{\delta \in \mathcal{D}} \mathrm{Var}\left[B(\Phi)|\delta\right]$, where

$$\mathrm{Var}\left[B(\Phi)|\delta\right] = \left(\frac{n}{r\sigma^2}\right)^2 \mathrm{Var}\left[\mathrm{tr}\left\{[FA(\delta)D(\delta) - I]'[FA(\delta)D(\delta) - I]\mathbf{\Phi}\mathbf{\Phi}'\right\}\right] .$$

As minimization of $\mathrm{Var}\left[B(\Phi)|\delta\right]$ does not necessarily lead to designs with small bias, VB-optimality is only recommended for use as a secondary criterion to discriminate between designs which are efficient under EB- or PB-optimality.

The task of finding designs analytically under the above criteria is mathematically intractable for even simple contamination functions. Hence a search approach is adopted and a modified Fedorov exchange algorithm (Cook and Nachtsheim (1980)) developed. An exchange algorithm swaps points between a candidate list of possible points and a design in an iterative procedure to find an optimal design. An exchange is kept if it results in an improvement in the chosen objective function. In the algorithm used here, for each exchange of a candidate and design point, the required properties of the bias distribution are approximated by a Monte Carlo simulation using samples of size s.

Implementation of an EB-optimal search requires, from (3.1), only one approximation of $E[\mathbf{\Phi}\mathbf{\Phi}']$ in each run of the algorithm. In outline, realizations $\phi_1, \ldots, \phi_s$ of the contamination are simulated and $E[\mathbf{\Phi}\mathbf{\Phi}']$ is approximated by $\sum_{i=1}^{s} \phi_i\phi_i'/n$. At each exchange of design and candidate points, a comparison is made of the values of the objective function for the current and potential designs using the single simulation. This leads to a computationally efficient search and the elimination of Monte Carlo error within each run of the algorithm.

By contrast, in order to implement the $PB(p)$- and VB-optimality criteria, a separate approximation of the appropriate bias property is required for every swap of a design and candidate point. Large values of s are then necessary to ensure sufficient accuracy of design comparisons within the algorithm. This leads to far more computational effort than is required for EB-optimality. Thus, these criteria are most practically used to choose a design from a shortlist of EB-optimal designs.

In the examples in this paper, designs are found under EB-optimality using $s = 100,000$ for each search and 10 tries of the algorithm with different random starting designs. This value of s was found to give accurate approximations of $E[\mathbf{\Phi}\mathbf{\Phi}']$. The design region was assumed to be the same as the region over which predictions are to be made. Hence the candidate list and the evaluation grid over which (3.1) is approximated were chosen to coincide. The algorithms used to generate and evaluate the designs in this paper are available from the author.

4. Application to Polynomial Spline Contamination

In many practical experiments, a factor may affect a response in a way that is not adequately described by a low order polynomial. For example, the response may have several local optima or display non-smooth behaviour. A model built from polynomial spline functions may then provide a better description of the data and more accurate predictions; see Wold (1974) and Frey (1993) for examples of spline modeling in chemistry, and Grove, Woods and Lewis (2004) for an automotive application.

A spline of degree d is a function that may take a different polynomial form, up to degree d, on each of a set of continuous intervals separated by points called *knots*. The polynomial pieces are constrained to be at least continuous at these knots. The splines considered here also have $d-1$ continuous derivatives at each knot and are known as *maximally smooth*. Bases for polynomial splines include the truncated power and B-spline bases. We consider models built from the truncated power basis, which have the advantage that extra knots can be incorporated easily through the addition of terms to the model. For models with known knots, Draper, Guttman and Lipow (1977) found all-bias designs which protect against known higher order terms.

The truncated power representation of a maximally smooth polynomial spline regression model for a single factor is obtained by the substitution into equation (1.1) of

$$f(x) = \sum_{i=0}^{d} \beta_i x^i + \sum_{j=1}^{l} \beta_{d+j} (x - \xi_j)_+^d \,, \tag{4.1}$$

where the function $(a)_+$ is defined as a if $a > 0$ and zero otherwise, and ξ_j $(j = 1, \ldots, l)$ are the knots. When the ξ_j are assumed known, then (4.1) is linear

in the unknown parameters, β_i, and linear model inference and design search techniques can be applied. If there is uncertainty in the number and location of the knots, then a more complicated design problem results.

Suppose that the locations of certain knots are known, for example, from a pilot experiment or scientific knowledge, but that there is uncertainty in the numbers and locations of possible additional knots. Suppose also that the aim of the experiment is to make inferences using model (4.1) in the presence of possible bias resulting from omitting the additional knots. We may then select designs using the criteria of Section 3.

In order to model our uncertainty we assume that the contamination consists of a random number of extra knots in random locations, so that each realization $\phi(x)$ has the form

$$\phi(x) = \sum_{i=1}^{k} \gamma_i (x - \lambda_i)_+^d \,.$$

We assume also that the number of additional knots, k, is a realization of a random variable K and that, conditional on $K = k$, the locations λ_i and coefficients γ_i are realizations of random variables Λ_i and Γ_i respectively $(i = 1, \ldots, k)$. Then

$$\Phi(x) = \sum_{i=1}^{K} \Gamma_i (x - \Lambda_i)^d \,. \tag{4.2}$$

Prior distributions may be placed on K, Λ_i and Γ_i using any available information on the contamination to suggest values for their respective means μ_k, μ_l, μ_g and variances σ_k^2, σ_l^2, σ_g^2. A prior distribution for $B(\Phi)$ can then be defined and the criteria of Section 3 applied.

Two examples are presented below where it is assumed that K follows a Poisson(μ_k) distribution and, for given $K = k$, the Γ_j are independent and identically distributed $N(\mu_g, \sigma_g^2)$ $(j = 1, \ldots, k)$. For each example, the factor x takes values in $[-1, 1]$. In Example 1, each Λ_i follows a Uniform(l_1, l_2) distribution and in Example 2, each Λ_i follows a Beta(u, v) distribution defined on the interval $[-1, 1]$. Also Λ_i, Γ_j $(i, j = 1, \ldots, k)$ are assumed to be mutually independent.

Example 1. Consider a quadratic polynomial assumed model, $d = 2$, $l = 0$. Four-point near EB-optimal designs were found from 10 runs of the algorithm for the values of μ_k, μ_g and σ_g^2 shown in Table 1 and with $l_1 = -1$ and $l_2 = 1$, so that additional knots are equally likely to occur in any subinterval of the range of x. The same design $\{-0.85, -0.35, 0.35, 0.85\}$ was obtained for each set of parameter values with approximate expected bias given in the table. This design has far lower expected bias than the D-optimal design $\{-1, 0, 0, 1\}$, between 62% and 65% smaller. The EB-optimal design also has much better performance for $n = 4$ points than the uniformly spread design $\{-1, -0.33, 0.33, 1\}$, with

expected bias between 49% and 54% smaller. As n increases, a uniformly spread design becomes increasingly competitive; for example, there is little difference in expected bias between the EB-optimal and the uniformly spread design for $n = 11$.

Table 1. The approximate expected bias for values of μ_k, μ_g and σ_g^2 from Example 1.

	Approx. expected bias		
μ_k, μ_g, σ_g^2	EB-optimal	Uniform spread	D-optimal
2, 0, 1	0.01	0.02	0.03
2, 10, 100	3.26	6.65	8.79
15, 0, 100	7.91	15.46	20.80
15, 10, 1	72.99	157.41	206.34

As the sub-intervals in which additional knots may occur are varied, different EB-optimal designs are obtained. Figure 1 shows designs for five different sets of l_1, l_2 values for $\mu_k = 15$, $\mu_g = 0$ and $\sigma_g^2 = 1$. All the designs have four distinct support points, two more than the D-optimal design for this assumed model, and the points become less equally-spaced as the l_1, l_2 values move away from -1, 1 respectively.

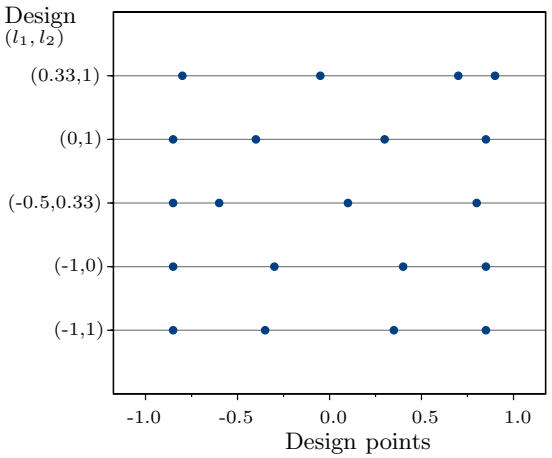

Figure 1. Five near EB-optimal designs, each found for different l_1 and l_2 values for Example 1 with $\mu_k = 15$, $\mu_g = 0$ and $\sigma_g^2 = 1$.

There can be many near-optimal designs for the random bias problem. Table 2 shows the designs found for 10 runs of the algorithm when $\mu_k = 2$, $\mu_g = 10$, $\sigma_g^2 = 1$, $l_1 = 0$ and $l_2 = 0.333$. There are eight distinct designs with some con-

siderable differences between their points. However, the maximum difference in approximate expected bias between these designs is only 2.5% (not shown). This difference, although small, demonstrates that the designs found are near, rather than exactly, optimal.

For this example, the size of the Monte Carlo error is negligible compared with the size of the expected bias. For assumed models with smaller expected bias, such as higher order polynomials or splines with many knots, larger simulations may be needed in order to discriminate between designs from different runs of the algorithm.

Table 2. Near EB-optimal designs found from ten runs of the exchange algorithm from Example 1, with approximate variance, median and the 95th bias percentile, together with the near $PB(0.95)$- and VB-optimal designs.

Design	Var (B)	Median	95th percentile
EB-optimal			
-0.85, -0.55, 0.15, 0.80	76.751	4.254	23.799
-0.80, -0.15, 0.60, 0.90	80.865	4.326	26.227
-0.85, -0.55, 0.15, 0.80	76.751	4.254	23.799
-0.80, -0.05, 0.60, 0.85	76.720	4.217	25.858
-0.75, -0.05, 0.15, 0.80	80.226	4.226	24.484
-0.75, 0.00, 0.70, 0.85	80.706	4.242	24.447
-0.80, -0.70, 0.05, 0.80	82.092	4.310	24.299
-0.85, -0.65, 0.70, 0.85	82.818	4.306	25.071
-0.75, 0.00, 0.70, 0.85	80.706	4.242	24.447
-0.75, 0.05, 0.75, 0.80	78.978	4.331	24.515
$PB(0.95)$-optimal			
-0.75, 0.05, 0.10, 0.80	77.582	4.285	23.770
VB-optimal			
-0.75, -0.05, 0.65, 0.80	75.912	4.250	25.011

Table 2 also gives the values of the objective functions for VB and $PB(p)$ ($p = 0.5$ and $p = 0.95$) for each design and shows only small differences (up to 8%) across the designs. Also given are the best approximate VB- and $PB(0.95)$-optimal designs found from ten tries for the same parameter settings. These two designs have similar variance bias and percentile bias values to those of the near EB-optimal designs. We conclude from this, and other examples, that the extra computational effort required to find the VB and PB-optimal designs does not result in a substantial increase in design performance and leads to the recommendation in Section 3.

Example 2. Consider a linear spline assumed model, $d = 1$, $l = 2$, with $\xi_1 = -0.333$ and $\xi_2 = 0.333$. The locations of the additional knots in the contamination function are assumed to follow Beta(u, v) distributions. For fixed $u = 5$ and $v = 10$, the eight-point EB-optimal design $\{-0.9, -0.6, -0.4, -0.2, 0, 0.25, 0.3, 0.8\}$ was found for all four of the triplets $(\mu_k,\ \mu_g,\ \sigma_g^2)$ with values (2, 10, 1), (2, 0, 100), (15, 0, 1) and (15, 10, 100). Designs found under different combinations of u and v are shown in Figure 2. Under this more flexible assumed model, containing two knots, and with the knot locations following a Beta distribution, the EB-optimal designs are less sensitive to the location parameters than in Example 1.

The invariance of EB-optimal designs to μ_k, μ_g and σ_g^2 demonstrated in Examples 1 and 2 has also been observed for many other examples and is established for the case when $\mu_g = 0$ in the following section.

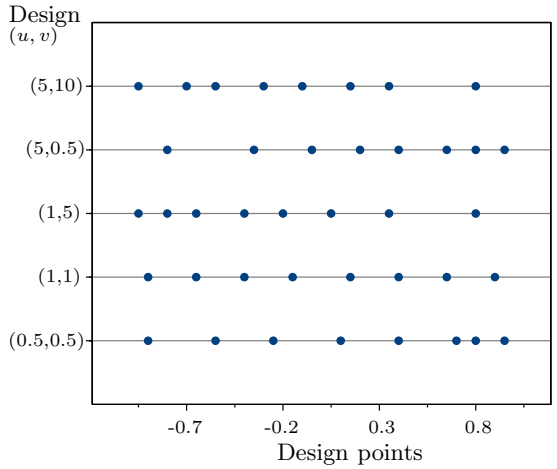

Figure 2. Near EB-optimal designs from Example 2 with $\mu_k = 2$, $\mu_g = 10$ and $\sigma_g^2 = 100$ for different u and v values.

5. A Theorem on *EB*-Optimal Designs

Suppose that $\mu_g = 0$, so that the expected value of each term in (4.2) is zero. Then the assumed model is correct on average but with any particular realization possibly being subject to non-zero contamination, see Notz (1989) and Dumouchel and Jones (1994).

The following lemma and theorem establish the invariance of an EB-optimal design to μ_k, σ_k^2 and σ_g^2 when $\mu_g = 0$. Proofs are given in the appendix.

Lemma 1. Assume (4.1) and (4.2), where K has a distribution with mean μ_k and variance σ_k^2 and, for given $K = k$, the knot locations Λ are identically and independently

distributed (i.i.d.). The corresponding coefficients Γ are also i.i.d. with mean μ_g and variance σ_g^2, and Λ and Γ are assumed independent. Then

$$E\left[\mathbf{\Phi}\mathbf{\Phi}'\right] = \mu_k(\sigma_g^2 + \mu_g^2)E^{(1)} + \mu_g^2\left(\sigma_k^2 + \mu_k(\mu_k - 1)\right)E^{(2)}\,, \tag{5.1}$$

where $\mathbf{\Phi} = (\Phi(x_1), \ldots, \Phi(x_r))'$, and $E^{(1)}$ and $E^{(2)}$ are $r \times r$ matrices with respective (i,j)th entries

$$E_{ij}^{(1)} = E\left[(x_i - \Lambda)_+^d (x_j - \Lambda)_+^d\right] \text{ and } E_{ij}^{(2)} = E\left[(x_i - \Lambda)_+^d\right] E\left[(x_j - \Lambda)_+^d\right]\,.$$

Theorem 1. *Under the assumptions of Lemma* 1*, if* $\mu_g = 0$ *then*

(i) $E[B(\Phi)|\delta] = n\mu_k\sigma_g^2 \mathrm{tr}\{[F(X'X)^{-1}X'D - I]'[F(X'X)^{-1}X'D - I]E^{(1)}\}/(r\sigma^2)$,

(ii) *the expected bias optimal design is invariant to* μ_k*,* σ_k^2 *and* σ_g^2*.*

The use of this theorem increases the computational efficiency of the design search by reducing the amount of simulation required to approximate the expected bias. Empirical evidence, such as that from Examples 1 and 2, suggests that even when $\mu_g \neq 0$, EB-optimal designs are unaffected by the values of μ_k, σ_k^2, μ_g and σ_g^2, and leads to the conjecture that Theorem 1 (ii) holds when $\mu_g \neq 0$.

6. Extension of Criteria to Expected Mean Squared Error

Consider the expected mean squared error under a design $\delta \in \mathcal{D}$:

$$E(MSE|\delta) = V(\delta) + E\left[B(\Phi)|\delta\right]\,, \tag{6.1}$$

where $V(\delta) = (n/r)\mathrm{tr}\left\{F'(X'X)^{-1}F\right\}$, with F defined as in Section 2. A design that minimizes $V(\delta)$ is called V-optimal. A design $\delta^* \in \mathcal{D}$ that minimizes (6.1) is *expected mean squared error optimal* ($EMSE$-optimal) over the set $\mathcal{D}$. Near optimal designs may be found using the methods described in Section 3; see also Allen, Yu and Schmitz (2003).

Under spline contamination (4.2), with $\mu_g = 0$, it follows from Theorem 1 that

$$E(MSE|\delta) = \frac{n}{r}\mathrm{tr}\left\{F'(X'X)^{-1}F + \frac{\mu_k\sigma_g^2}{\sigma^2}[FAD - I]'[FAD - I]E^{(1)}\right\}\,.$$

For a given assumed model and known distributions for the additional knot locations, the ratio of $\mu_k\sigma_g^2$ to σ^2 controls the trade-off between the variance error and the bias in design selection. Large values of this ratio correspond to a larger expected number of additional knots and/or a more diffuse prior for Γ_i. Hence an $EMSE$-optimal design depends on this ratio, as demonstrated in the following example.

Example 3. Consider (4.1) with $l = 0$, $d = 2$, and (4.2), where $\Lambda_i \sim U(-0.2, 0.2)$ and $\mu_g = 0$. Table 3 gives four-point approximate $EMSE$-optimal designs for different values of the ratio $R = \mu_k\sigma_g^2/\sigma^2$. As $R \to \infty$, the $EMSE$-optimal design approaches the EB-optimal design $\{-0.8, -0.2, 0.2, 0.8\}$. As R decreases, the design points slowly shift

toward the V-optimal design $\{-1, 0, 0, 1\}$ but the V-optimal design is not achieved until the square root of $V(\delta)$ is more than 3.8 times the size of the square root of $E(B(\Phi)|\delta)$. At this point, the bias would be difficult to detect and so the assumed model would be a good approximation. The limiting EB- and V-optimal designs are determined by the distribution of the additional knot locations and the assumed model, respectively. The table shows that a series of compromise designs is found between these two extremes.

Table 3. Approximate optimal expected mean squared error designs for different values of the ratio R. $*$ (†) denotes the EB-optimal (V-optimal) design.

$\frac{\mu_k \sigma_g^2}{\sigma^2}$	$\sqrt{\frac{V(\delta)}{E(B[\Phi]\|\delta]}}$	Design			
$\to \infty$	$\to 0$	-0.8,	-0.2,	0.2,	0.8^*
213	1.0	-0.85,	-0.25,	0.25,	0.85
90	1.4	-0.9,	-0.3,	0.3,	0.9
25	2.3	-0.95,	-0.25,	0.25,	0.95
10	2.8	-1,	-0.15,	0.15,	1
5	3.8	-1,	0,	0,	$1^\dagger$

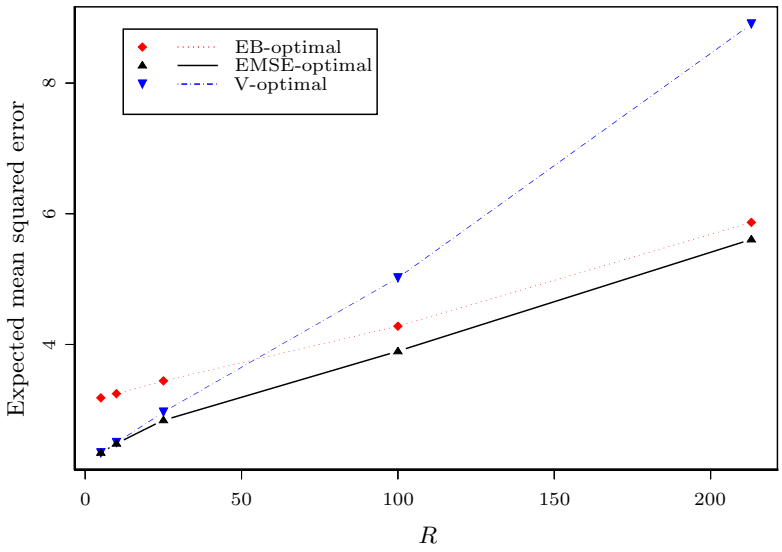

Figure 3. Approximate expected mean squared error for the $EMSE$-, EB- and V-optimal designs of Example 3 for a selection of R values.

Figure 3 shows that, although for small values of R the V-optimal design has smaller expected mean squared error than the EB-optimal design, the expected mean squared error for the V-optimal design increases rapidly with R. For the

EB-optimal design, this increase in $E[MSE]$ is slower. The expected mean squared error for the EB-optimal design is smaller than for the V-optimal design for $R \geq 50$ and tends towards the value for the $EMSE$-optimal design as $R \to \infty$. The advantage of the $EMSE$-optimal design is clear: for *any* value of R, it has the lowest possible value for $E[MSE]$.

7. Discussion

For the EB-criterion, a computationally efficient design search is possible which produces designs that perform well under the $PB(p)$ and VB criteria. The designs have more distinct points, more evenly distributed across the design region, than designs obtained from variance-based criteria. These properties can make the designs more attractive for many practical experiments.

The criteria are particularly appropriate for polynomial spline models where there will generally be uncertainty in the number and location of knots prior to the experiment. It is then useful to have designs which offer some protection against additional unknown knots. The robustness of the EB-optimal designs to the number of extra knots and the distributions of the contamination parameters speeds the design search and increases the applicability of the designs.

Acknowledgement

The author is grateful for guidance from his Ph.D. supervisor S. M. Lewis (University of Southampton) and helpful discussions with A. H. Welsh (Australian National University), A. M. Dean (Ohio-State University) and J. G. Frey (University of Southampton). The paper benefitted from the comments of two referees. The work was funded by the EPSRC grant GR/R67729.

Appendix

Proof of Lemma 1. From (4.2), it follows that

$$\begin{aligned}\Phi(x_i)\Phi(x_j) &= \left[\sum_{s=1}^{K}\Gamma_s(x_i-\Lambda_s)_+^d\right]\left[\sum_{t=1}^{K}\Gamma_t(x_j-\Lambda_t)_+^d\right] \\ &= \sum_{h=1}^{K}\Gamma_h^2(x_i-\Lambda_h)_+^d(x_j-\Lambda_h)_+^d + \sum_{\substack{s=1\\ s\neq t}}^{k}\sum_{t=1}^{k}\Gamma_s\Gamma_t(x_i-\Lambda_s)_+^d(x_j-\Lambda_t)_+^d\,,\end{aligned}$$

and hence

$$E\left[\Phi(x_i)\Phi(x_j)\right] = \sum_{k=0}^{\infty} E\left[\Phi(x_i)\Phi(x_j)|K=k\right]P(K=k)\,. \tag{A.1}$$

Further, by independence of Λ_s, Λ_t, Γ_s and Γ_t, it can be shown that

$$E\left[\Phi(x_i)\Phi(x_j)|K=k\right] = k(\sigma_g^2+\mu_g^2)E\left[(x_i-\Lambda)_+^d(x_j-\Lambda)_+^d\right] + k(k-1)\mu_g^2 E\left[(x_i-\Lambda)_+^d\right]E\left[(x_j-\Lambda)_+^d\right], \quad \text{(A.2)}$$

as Λ_s $(s=1,\ldots,k)$ are identically distributed. Substitution of (A.2) into (A.1) gives

$$\begin{aligned} E\left[\Phi(x_i)\Phi(x_j)\right] &= (\sigma_g^2+\mu_g^2)E\left[(x_i-\Lambda)_+^d(x_j-\Lambda)_+^d\right]\sum_{k=0}^{\infty}kP(K=k) \\ &\quad +\mu_g^2 E\left[(x_i-\Lambda)_+^d\right]E\left[(x_j-\Lambda)_+^d\right]\sum_{k=0}^{\infty}k(k-1)P(K=k) \\ &= \mu_k(\sigma_g^2+\mu_g^2)E\left[(x_i-\Lambda)_+^d(x_j-\Lambda)_+^d\right] \\ &\quad +\mu_g^2\left(\sigma_k^2+\mu_k(\mu_k-1)\right)E\left[(x_i-\Lambda)_+^d\right]E\left[(x_j-\Lambda)_+^d\right]. \end{aligned}$$

Proof of Theorem 1. From (3.1),

$$E[B(\Phi)|\delta] = \frac{n}{r\sigma^2}\mathrm{tr}\left\{[F(X'X)^{-1}X'D-I]'[F(X'X)^{-1}X'D-I]E\left[\mathbf{\Phi}\mathbf{\Phi}'\right]\right\},$$

where $\mathbf{\Phi}\mathbf{\Phi}'$ does not depend on $\delta\in\mathcal{D}$. An application of Lemma 1 with $E[\Gamma]=\mu_g=0$, gives

$$E\left[B(\Phi)|\delta\right] = \mu_k\sigma_g^2\frac{n}{r\sigma^2}\mathrm{tr}\left\{[F(X'X)^{-1}X'D-I]'[F(X'X)^{-1}X'D-I]E^{(1)}\right\},$$

where $E^{(1)}$ and $E^{(2)}$ are defined in the statement of the lemma. This expression is minimized by a choice of design δ that minimizes

$$\frac{n}{r\sigma^2}\mathrm{tr}\left\{[F(X'X)^{-1}X'D-I]'[F(X'X)^{-1}X'D-I]E^{(1)}\right\}$$

regardless of the values of μ_k, σ_k^2 and σ_g^2.

References

Allen, T., Yu, L. and Schmitz, J. (2003). An experimental design criterion for minimizing meta-model prediction errors applied to die casting process design. *J. Roy. Statist. Soc. Ser. C.* **52**, 103-117.

Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54**, 622-654.

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

Draper, N. R., Guttman, I. and Lipow, P. (1977). All-bias designs for spline functions joined at the axes. *J. Amer. Statist. Assoc.* **72**, 424-429.

Dumouchel, W. and Jones, B. (1994). A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model. *Technometrics* **36**, 37-47.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing.* Marcel Dekker, New York.

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

Frey, J. G. (1993). Analysis of time of flight data: constraints and error bars. *Mol. Phys.* **79**, 1287-1303.

Grove, D. M., Woods, D. C. and Lewis, S. M. (2004). Multifactor B-spline mixed models in designed experiments for the engine mapping problem. *J. Qual. Technol.* **36**, 380-391.

Heo, G., Schmuland, B. and Wiens, D. P. (2001). Restricted minimax robust designs for misspecified linear models. *Canad. J. Statist.* **29**, 117-128.

Montepiedra, G. and Fedorov, V. V. (1997). Minimum bias designs with constraints. *J. Statist. Plann. Inference* **63**, 97-111.

Myers, R. H. and Montgomery, D. C. (2002). *Response Surface Methodology.* Wiley, New York.

Notz, W. I. (1989). Optimal designs for regression models with possible bias. *J. Statist. Plann. Inference* **22**, 43-54.

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

Wiens, D. P. and Zhou, J. (1997). Robust designs based on the infinitesimal approach. *J. Amer. Statist. Assoc.* **92**, 1503-1511.

Wold, S. (1974). Spline functions in data analysis. *Technometrics* **16**, 1-11.

Southampton Statistical Sciences Research Institute, School of Mathematics, University of Southampton, Southampton, SO17 1BJ, U.K.

E-mail: D.C.Woods@maths.soton.ac.uk

(Received December 2003; accepted June 2004)