Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 1339–1357
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Robust designs for one-point extrapolation

Douglas P. Wiens[a], Xiaojian Xu[b,*]

[a]Department of Mathematical and Statistical Sciences, University of Alberta, Canada
[b]Department of Mathematics, Brock University, St. Catharines, Ont., Canada L2S 3A1

Received 25 May 2005; received in revised form 3 March 2007; accepted 15 June 2007
Available online 3 July 2007

Abstract

We consider the construction of designs for the extrapolation of a regression response to one point outside of the design space. The response function is an only approximately known function of a specified linear function. As well, we allow for variance heterogeneity. We find minimax designs and corresponding optimal regression weights in the context of the following problems: (P1) for nonlinear least squares estimation with homoscedasticity, determine a design to minimize the maximum value of the mean squared extrapolation error (MSEE), with the maximum being evaluated over the possible departures from the response function; (P2) for nonlinear least squares estimation with heteroscedasticity, determine a design to minimize the maximum value of MSEE, with the maximum being evaluated over both types of departures; (P3) for nonlinear weighted least squares estimation, determine both weights and a design to minimize the maximum MSEE; (P4) choose weights and design points to minimize the maximum MSEE, subject to a side condition of unbiasedness. Solutions to (P1)–(P4) are given in complete generality. Numerical comparisons indicate that our designs and weights perform well in combining robustness and efficiency. Applications to accelerated life testing are highlighted.

MSC: primary 62K05; 62F35; secondary 62J12

Keywords: Accelerated life testing; Generalized linear model; Nonlinear least squares; Regression design

1. Introduction

In this article, we study the construction of designs for the extrapolation of regression responses to one point outside of the design space. Such "one-point extrapolation" designs are of interest in problems of accelerated life testing (ALT), in which products are typically tested at unusual stress levels, with the results then extrapolated to a lower stress level anticipated in practice. Our model is somewhat similar to a generalized linear model, in that the response fitted by the experimenter is a function of a linear function of unknown parameters and known regressors. Our designs are robust in that we allow both for imprecision in the specification of the response, and for possible heteroscedasticity.

Robust designs for extrapolation of a, possibly misspecified, *linear* response were obtained by Fang and Wiens (1999); see also the references therein, in particular Dette and Wong (1996), Draper and Herzberg (1973), Huang and Studden (1988), Huber (1975) and Spruill (1984). The current work goes beyond Fang and Wiens (1999) in two

* Corresponding author.
E-mail addresses: doug.wiens@ualberta.ca (D.P. Wiens), xxu@brocku.ca (X. Xu).

0378-3758/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2007.06.020

ways—in the move to a generalized linear response as described above, and in our emphasis on extrapolation to a single point, thus allowing for more explicit and applicable results than were previously possible.

For nonlinear regression, Ford et al. (1989) present various static and sequential designs for nonlinear models without the consideration of model uncertainty. Sinha and Wiens (2002) have employed notions of robustness in the construction of sequential designs for the nonlinear model. In many ALT applications however, sequential designs are not feasible (Ford et al., 1989), hence our focus in this article on static designs.

Fang and Wiens (1999) point out that "Extrapolation to regions outside of that in which observations are taken is, of course, an inherently risky procedure and is made even more so by an over-reliance on stringent model assumptions." With this in mind, we shall depart rather broadly from the usual generalized linear response models:

1. The response is taken to be an approximately known function of a linear function of known regressors and unknown parameters:
$$E(Y|\mathbf{x}) = h(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x})) + n^{-1/2} f(\mathbf{x})$$
for p regressors $\mathbf{z}(\mathbf{x}) = (z_1(\mathbf{x}), z_2(\mathbf{x}), \ldots, z_p(\mathbf{x}))^{\mathrm{T}}$, depending on a q-dimensional vector $\mathbf{x}$ of independent variables. The function h is strictly monotonic, with a bounded second derivative. We assume that $\|\mathbf{z}(x)\|$ is bounded on S. The response contaminant f represents uncertainty about the exact nature of the regression response and is unknown and arbitrary, subject to certain restrictions. We estimate $\boldsymbol{\theta}$ but not f; this leads to possibly biased extrapolations $\hat{Y}(\mathbf{x}) = h(\hat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z}(\mathbf{x}))$ of $E(Y|\mathbf{x})$. The factor $n^{-1/2}$ is necessary for a sensible asymptotic treatment. It ensures that losses due to bias remain of the same asymptotic order as those due to variance, and is analogous to the requirement of contiguity in the asymptotic theory of hypothesis testing. As in this analogous testing situation it has no effect in finite samples, since the factor can be absorbed into the alternative, i.e. into f.
2. The experimenter takes n uncorrelated observations $Y_i = Y(\mathbf{x}_i)$, with $\mathbf{x}_i$ freely chosen from a design space S. Our goal is to choose these design points from S in an optimal manner in order to extrapolate the estimates of $E(Y|\mathbf{x})$ to $\mathbf{x}_0$.
3. The observations Y_i are possibly heteroscedastic, with $VAR\{Y(\mathbf{x}_i)\} = \sigma^2 g(\mathbf{x}_i)$ for a function g satisfying conditions given below.

We estimate $\boldsymbol{\theta}$ by least squares (LS), possibly weighted with weights $w(\mathbf{x})$. Our loss function is n times the mean-squared error of $\hat{Y}(\mathbf{x}_0)$ in estimating $E(Y|\mathbf{x}_0)$. This depends on the design measure $\xi = n^{-1}\Sigma_{i=1}^n \delta_{\mathbf{x}_i}$ as well as on w, f and g:
$$\mathrm{MSEE}(f, g, w, \xi) = nE\{[\hat{Y}(\mathbf{x}_0) - E(Y|\mathbf{x}_0)]^2\}.$$

We denote unweighted least squares by $w = \mathbf{1}$, and homogeneous variances by $g = \mathbf{1}$. The following problems will be addressed:

(P1) For ordinary least squares (OLS) estimation under homoscedasticity, determine designs to minimize the maximum value, over f, of $\mathrm{MSEE}(f, \mathbf{1}, \mathbf{1}, \xi)$.
(P2) For OLS estimation under possible heteroscedasticity, determine designs to minimize the maximum value, over f and g, of $\mathrm{MSEE}(f, g, \mathbf{1}, \xi)$.
(P3) For weighted least squares (WLS) estimation, determine designs and weights to minimize the maximum value, over f and g, of $\mathrm{MSEE}(f, g, w, \xi)$.
(P4) Choose weights and design points to minimize $\max_{f,g} \mathrm{MSEE}(f, g, w, \xi)$, subject to a side condition of unbiasedness.

The overall message to be taken from this work is that the experimenter must be extremely wary of an over-reliance on an assumed model from which to extrapolate predictions. The performance of a design which is optimal, in the sense of minimizing variance at the assumed model, typically deteriorates rapidly as the assumed and true models diverge, and as bias is included in the loss. Stark illustrations of this are given in Example 2 of Section 5, where we compare the robust, classically optimal and naive uniform designs for straight line and quadratic regression. Further in Section 7, we demonstrate that for minimization of variance alone the designs of Hoel and Levine (1964) are optimal, and significantly better than their competitors if the extrapolation point is not far from the design space. When bias is included, however, the Hoel–Levine designs can perform very poorly, with an *MSEE* which is unbounded in the presence of the types of contamination considered here.

The rest of this article is organized as follows. The designs for P1 are provided in Section 4. The designs and weights which constitute solutions to problems P2 and P3 are given in Section 5. Those for P4 are given in Section 6. Some mathematical preliminaries are detailed in Section 2. The maximization part of the design construction is provided in Section 3. Comparisons of these designs are presented in Section 7. All proofs are in the Appendix.

2. Preliminaries and notation

We define the "target" parameter $\boldsymbol{\theta}_0$ to be that which produces the best agreement, in the L_2-sense, between $h(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}))$ and $E(Y|\mathbf{x})$:

$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}} \left\{ \int_S [h(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x})) - E(Y|\mathbf{x})]^2 \,\mathrm{d}\mathbf{x} \right\}.$$

We assume an open parameter space, so that with

$$f_n(\mathbf{x}) = \sqrt{n}[E(Y|\mathbf{x}) - h(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x}))]$$

and

$$\tilde{\mathbf{z}}(\mathbf{x}) = h'(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x}))\mathbf{z}(\mathbf{x})$$

we have $\int_S \tilde{\mathbf{z}}(\mathbf{x}) f_n(\mathbf{x}) \,\mathrm{d}\mathbf{x} = \mathbf{0}$. Where possible we drop the subscript on f.

We shall assume that $f_n = f$ is an unknown member of the class

$$\mathscr{F} = \left\{ f \,\middle|\, \int_S f^2\ (\mathbf{x}) \,\mathrm{d}\mathbf{x} \leqslant \eta_S^2 < \infty, |f(\mathbf{x}_0)| \leqslant \eta_T < \infty, \int_S \tilde{\mathbf{z}}(\mathbf{x}) f(\mathbf{x}) \,\mathrm{d}\mathbf{x} = \mathbf{0} \right\},$$

for positive constants η_S, η_T. Our theory makes no assumptions, and imposes no requirements, on the behaviour of members of $\mathscr{F}$ off of $S \cup \{\mathbf{x}_0\}$. One might then, without loss of generality, require that they be bounded, or continuous, etc. in this complementary region.

The departure from homogeneity of variances is measured by $g(\mathbf{x})$, which is assumed to be an unknown member of the class

$$\mathscr{G} = \left\{ g \,\middle|\, \int_S g^2\ (\mathbf{x}) \,\mathrm{d}\mathbf{x} \leqslant \Omega^{-1} := \int_S \mathrm{d}\mathbf{x} < \infty \right\}. \tag{1}$$

The condition in (1) is equivalent to defining $\sigma^2 = \sup_g \{\int_S var^2[\varepsilon(\mathbf{x})]\Omega \,\mathrm{d}\mathbf{x}\}^{1/2}$.

To ensure the nonsingularity of a number of relevant matrices, we assume that the regressors and design space satisfy

(A) For each $\mathbf{a} \neq \mathbf{0}$, the set $\{\mathbf{x} \in S : \mathbf{a}^{\mathrm{T}}\tilde{\mathbf{z}}(\mathbf{x}) = 0\}$ has Lebesgue measure zero.

We propose to estimate $\boldsymbol{\theta}_0$ using LS to fit $E(\hat{Y}|\mathbf{x}) = h(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x}))$ with nonnegative weights $w(\mathbf{x})$.

We make use of the following matrices and vectors:

$$\mathbf{A}_S = \int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) \,\mathrm{d}\mathbf{x}, \quad \mathbf{A}_T = \tilde{\mathbf{z}}(\mathbf{x}_0)\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}_0),$$
$$\mathbf{B} = \int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) w(\mathbf{x})\xi(\mathrm{d}\mathbf{x}), \quad \mathbf{D} = \int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) w^2(\mathbf{x}) g(\mathbf{x})\xi(\mathrm{d}\mathbf{x}),$$
$$\mathbf{b}_{f,S} = \int_S \tilde{\mathbf{z}}(\mathbf{x}) f(\mathbf{x}) w(\mathbf{x})\xi(\mathrm{d}\mathbf{x}), \quad \mathbf{b}_{f,T} = \tilde{\mathbf{z}}(\mathbf{x}_0) f(\mathbf{x}_0).$$

It follows from (A) that $\mathbf{A}_S$ is nonsingular. The LS estimator of $\boldsymbol{\theta}_0$ is

$$\hat{\boldsymbol{\theta}} = \arg\min \sum_{i=1}^{n} [Y_i - h(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}))]^2 w(\mathbf{x}_i)$$

and satisfies $\sum_{i=1}^{n}\dot{\boldsymbol{\phi}}_i(\hat{\boldsymbol{\theta}})=0$ for

$$\dot{\boldsymbol{\phi}}_i(\boldsymbol{\theta})=[Y_i-h(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))][h'(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))]w(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i).$$

In addition, the Hessian $\ddot{\Phi}(\boldsymbol{\theta})$ is given by

$$\sum_{i=1}^{n}\ddot{\phi}_i(\boldsymbol{\theta})=\sum_{i=1}^{n}[Y_i-h(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))][h''(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))]w(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i)\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)-\sum_{i=1}^{n}[h(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))]^2w(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i)\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i).$$

The information matrix is

$$\mathscr{I}(\boldsymbol{\theta}_0)=\lim_{n\to\infty}E\left(-\frac{1}{n}\ddot{\Phi}(\boldsymbol{\theta}_0)\right)=\mathbf{B},$$

since

$$\begin{aligned}&E\left\{\frac{1}{n}\sum_{i=1}^{n}[Y_i-h(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))][h''(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))]w(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i)\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\right\}\\&\quad=n^{-1/2}\cdot\frac{1}{n}\sum_{i=1}^{n}f(\mathbf{x}_i)[h''(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))]w(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i)\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\end{aligned}$$

is $\mathrm{O}(n^{-1/2})$ by virtue of our assumptions on f, h and $\mathbf{z}$.

By Taylor's Theorem,

$$\mathbf{0}=\sum_{i=1}^{n}\dot{\boldsymbol{\phi}}_i(\hat{\boldsymbol{\theta}})=\sum_{i=1}^{n}\{\dot{\boldsymbol{\phi}}_i(\boldsymbol{\theta}_0)+\ddot{\boldsymbol{\phi}}_i(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\},$$

where $\tilde{\boldsymbol{\theta}}$ lies between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Then

$$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=\left(-\frac{1}{n}\sum_{i=1}^{n}\ddot{\phi}_i(\tilde{\boldsymbol{\theta}})\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{\boldsymbol{\phi}}_i(\boldsymbol{\theta}_0)\right).$$

Note that $n^{-1/2}\sum_{i=1}^{n}\dot{\boldsymbol{\phi}}_i(\boldsymbol{\theta}_0)$ is asymptotically normal, with asymptotic mean $\mathbf{b}_{f,S}$ and covariance

$$COV\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{\boldsymbol{\phi}}_i(\boldsymbol{\theta}_0)\right]=\frac{1}{n}\sum_{i=1}^{n}[h(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))]^2\sigma^2g(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i)\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)w^2(\mathbf{x}_i)=\sigma^2\mathbf{D}.$$

As at (Seber and Wild, 2003, Section 12.2), it follows that the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ is

$$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\sim\mathrm{AN}(\mathbf{B}^{-1}\mathbf{b}_{f,S},\sigma^2\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}),$$

and then by the delta method,

$$\sqrt{n}(h(\hat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_0))-h(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x}_0)))\sim\mathrm{AN}(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}_0)B^{-1}\mathbf{b}_{f,S},\sigma^2\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}_0)B^{-1}\mathbf{D}\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)).$$

The loss function *MSEE* splits into terms due to (squared) extrapolation bias (EB) and extrapolation variance (EV):

$$\begin{aligned}\mathrm{MSEE}(f,g,w,\xi)&=nE\{[\hat{Y}(\mathbf{x}_0)-E(Y|\mathbf{x}_0)]^2\}\\&=nE\left\{[h(\hat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_0))-h(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x}_0))-\frac{1}{\sqrt{n}}f(\mathbf{x}_0)]^2\right\}\\&=\mathrm{EB}(f,w,\xi)+\mathrm{EV}(g,w,\xi),\end{aligned}$$

where the squared EB and EV are

$$\mathrm{EB}(f,w,\xi)=\{\sqrt{n}E[h(\hat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_0))-h(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x}_0))]-f(\mathbf{x}_0)\}^2$$

and

$$\mathrm{EV}(g,w,\xi)=nVAR(\hat{Y}(\mathbf{x}_0))=nVAR(h(\hat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_0))).$$

Asymptotically,

$$\mathrm{EB}(f,w,\xi)=\mathbf{b}_{f,S}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\mathbf{b}_{f,S}-2\mathbf{b}_{f,T}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{b}_{f,S}+f^2(\mathbf{x}_0), \tag{2}$$
$$\mathrm{EV}(g,w,\xi)=\sigma^2\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}_0)\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)=\sigma^2\mathrm{tr}\mathbf{A}_T\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}. \tag{3}$$

We have defined ξ to be a discrete measure, with atoms of size n^{-1} at the design points (possibly repeated). We now adopt the viewpoint of approximate design theory and allow ξ to be any probability measure on S. One reason for this is that as at Lemma 1 of Wiens (1992), the class $\mathscr{F}$ is so broad that only absolutely continuous measures ξ can have finite maximum loss. Thus, let $k(\mathbf{x})$ be the density of ξ, and define $m(\mathbf{x})=k(\mathbf{x})w(\mathbf{x})$. Without loss of generality, we assume that the mean weight is $\int_S w(\mathbf{x})\xi(\mathrm{d}\mathbf{x})=1$. Then $m(\mathbf{x})$ is also a density on S which satisfies

$$\int_S \frac{m(\mathbf{x})}{w(\mathbf{x})}\,\mathrm{d}\mathbf{x}=1, \tag{4}$$

and we have

$$\mathbf{B}=\int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})m(\mathbf{x})\,\mathrm{d}\mathbf{x},$$
$$\mathbf{b}_{f,S}=\int_S \tilde{\mathbf{z}}(\mathbf{x})f(\mathbf{x})m(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

Remark 1. The requirement of absolute continuity excludes exact, implementable designs, and so approximations are necessary. We recall the statement in (Wiens, 1992 p. 355): "Our attitude is that an approximation to a design which is robust against more realistic alternatives is preferable to an exact solution in a neighbourhood which is unrealistically sparse." Various methods for implementing designs with continuous measures are discussed in Heo et al. (2001) and the references therein. As an example, a practical implementation for univariate x is to place the n design points at the quantiles $x_i=\xi^{-1}((i-1)/(n-1))$.

There has been some recent research into contamination classes which are sufficiently broad as to encompass realistic alternatives to fitted models, while admitting discrete optimal designs. Of note is the use of reproducing kernel Hilbert spaces, as in Yue and Hickernell (1999). It appears, however, that the theoretical and computational difficulties of this approach have so far limited its application to the few cases considered there. If the design space is finite, then exact designs can be obtained which are robust against a discrete version of the class $\mathscr{F}$—see Fang and Wiens (2000) for examples.

From the definitions of $\mathbf{B}$, $\mathbf{b}_{f,S}$ and $\mathbf{b}_{f,T}$, we notice that $\mathrm{EB}(f,w,\xi)$ depends on (w,ξ) only through m and $\mathrm{EV}(g,w,\xi)$ through m and w. Hence, we can optimize over m and w subject to (4) rather than over k and w. In the next four sections we exhibit solutions to P1–P4.

3. Maximization over $f \in F$ and $g \in G$

In this section, we exhibit the maxima of *MSEE*, for fixed functions $m(\mathbf{x})$ and $w(\mathbf{x})$. The minimizing m and w then constitute the solutions to P1–P4. The maxima are obtained in a manner very similar to that used in Fang and Wiens (1999), and so their derivations are omitted.

Define positive semidefinite matrices

$$\mathbf{K}=\int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})m^2(\mathbf{x})\,\mathrm{d}\mathbf{x},$$
$$\mathbf{G}=\mathbf{K}-\mathbf{B}\mathbf{A}_S^{-1}\mathbf{B}=\int_S[(m(\mathbf{x})\mathbf{I}-\mathbf{B}\mathbf{A}_S^{-1})\tilde{\mathbf{z}}(\mathbf{x})][(m(\mathbf{x})\mathbf{I}-\mathbf{B}\mathbf{A}_S^{-1})\tilde{\mathbf{z}}(\mathbf{x})]^{\mathrm{T}}\,\mathrm{d}\mathbf{x},$$

and constants $r_{T,S} = \eta_T/\eta_S$, reflecting the relative amounts of model response uncertainty in the extrapolation and design space, and $v = \sigma^2/\eta_S^2$, representing the relative importance of variance versus bias. In this notation, we have the following theorem.

Theorem 1. *The maximum squared EB is*

$$\sup_{f\in\mathscr{F}} \mathrm{EB}(f, m) = \eta_S^2(\sqrt{\lambda_m} + r_{T,S})^2,$$

where $\lambda_m = \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}_0)\mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)$. *The maximum is attained at*

$$f_m(\mathbf{x}) = \begin{cases} \eta_S \mathbf{z}^{\mathrm{T}}(\mathbf{x})\{m(\mathbf{x})\mathbf{I} - \mathbf{A}_S^{-1}\mathbf{B}\}\mathbf{a}_0, & \mathbf{x} \in S, \\ -\eta_T, & \mathbf{x} = \mathbf{x}_0, \end{cases}$$

where $\mathbf{a}_0 = \mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)/\sqrt{\lambda_m}$.

Remark 2. Recall Remark 1 above. It is of some interest to determine the bias which might arise in a discretized design. This bias is nonzero only for functions f placing mass at the design points and/or at $\mathbf{x}_0$, and is therefore bounded for uniformly bounded f. For a discrete design with probabilities ξ_i and regression weights w_i at the design points $\mathbf{x}_i$ ($i = 1, \ldots, N$) we calculate from (2) that the maximum squared EB, over the set of functions f with $\sum_{i=1}^N f^2(\mathbf{x}_i) = k^2$, is

$$\max_{\|\mathbf{f}\|=k} \mathrm{EB} = (\eta_T + k\|\mathbf{M}\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}^{\mathrm{T}}\mathbf{M}\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)\|)^2, \tag{5}$$

here we use the notation $m_i = \xi_i w_i$, $\mathbf{M} = \mathrm{diag}(m_1, \ldots, m_N)$, $\tilde{\mathbf{Z}} = (\tilde{\mathbf{z}}(\mathbf{x}_1) \cdots \tilde{\mathbf{z}}(\mathbf{x}_N))^{\mathrm{T}}$. This maximum EB is attained at $(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N))^{\mathrm{T}} \stackrel{\mathrm{def}}{=} \mathbf{f} = c\mathbf{M}\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}^{\mathrm{T}}\mathbf{M}\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)$, where c is chosen so that $\|\mathbf{f}\| = k$, and the sign of c is opposite to that of $f(\mathbf{x}_0)$. In Example 2 of Section 5, we illustrate our discretization methods, and compare the resulting designs with some common competitors, using as performance measures $\|\mathbf{M}\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}^{\mathrm{T}}\mathbf{M}\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)\|$, which is the only component of (5) which depends on the design, and an analogous discretization of $\mathrm{EV}(g, w, \xi)$. We find there that the discretized designs continue to enjoy favourable robustness properties with respect to bias, at a small cost in increased variance relative to the variance-minimizing designs. We do not present the discrete versions of the designs in other examples, since they tell much the same story as in Example 2.

We obtain Theorems 2 and 3 from Theorem 1. Theorem 2 gives the maximum MSEE under homoscedasticity while Theorem 3 gives this quantity under heteroscedasticity.

Theorem 2. *The maximum MSEE in problem P1 is*

$$\sup_{f\in\mathscr{F}} \mathrm{MSEE}(f, \mathbf{1}, \mathbf{1}, m) = \eta_S^2\{(\sqrt{\lambda_m} + r_{T,S})^2 + v\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}_0)\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)\}, \tag{6}$$

attained at f_m.

Theorem 3. *Define* $l_m(\mathbf{x}) = [\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)]^2$ *and* $\alpha_m = \int_S [l_m(\mathbf{x})m^2(\mathbf{x})]^{2/3}\,\mathrm{d}\mathbf{x}$. *Then the maximum MSEE in problems P2–P4 is*

$$\sup_{f\in\mathscr{F}, g\in\mathscr{G}} \mathrm{MSEE}(f, g, w, m) = \eta_S^2\left\{(\sqrt{\lambda_m} + r_{T,S})^2 + v\Omega^{-1/2}\left[\int_S \{w(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x})\}^2\,\mathrm{d}\mathbf{x}\right]^{1/2}\right\},$$

attained at f_m *and*

$$g_{m,w}(\mathbf{x}) \propto w(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x}).$$

The following theorem, whose proof is very similar to that of Theorem 2.(a) in Fang and Wiens (1999), gives the minimax weights for fixed $m(\mathbf{x})$.

Table 1
Numerical values for Example 1 with $S = [0, 1]$, $\theta_1 \in [0, 2]$, $r_{TS} = 1$, and $x_0 = 1.17$

v	a_1	a_2	a_3	a_5	θ_1^{LF}
0.5	0.130	0.421	−0.578	0.736	2
1	0.344	0.000224	−0.778	1.28	2
2	0.173	0.000286	−0.885	1.20	2

Theorem 4. *For fixed* $m(\mathbf{x})$ *the weights minimizing* $\sup_{f\in\mathscr{F},g\in\mathscr{G}} \mathrm{MSEE}(f, g, w, m)$ *subject to* (4) *are given by*

$$w_m(\mathbf{x}) = \alpha_m[l_m^2(\mathbf{x})m(\mathbf{x})]^{-1/3} I[m(\mathbf{x}) > 0].$$

Then $\min_w\{\sup_{f\in\mathscr{F},g\in\mathscr{G}} \mathrm{MSEE}(f, g, w, m)\} = \eta_S^2\ \{(\sqrt{\lambda_m} + r_{T,S})^2 + v\Omega^{-1/2}\alpha_m^{3/2}\}$.

4. Optimal designs with homoscedasticity: solution to P1

Problem P1 has become that of finding a density $m_*(\mathbf{x})$ which minimizes (6). The solution is given by Theorem 5, which reduces the problem to a $(2p + 1)$-dimensional numerical problem. The generality of our solution to P1, as well as those to P2 and P3, should be compared with the corresponding development in Fang and Wiens (1999). This generality, and the relative simplicity of the solutions, is made possible by our use of a one-point extrapolation region.

Theorem 5. *The density* $m_*(\mathbf{x})$ *minimizing* (6) *for OLS estimation under homoscedasticity is of the form*

$$m_*(\mathbf{x}) = \left[\frac{\mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\gamma}}{\mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}} + \frac{\lambda}{(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})^2}\right]^+,$$

where $(z)^+ = \max(z, 0)$. *The* $p \times 1$ *vectors* $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, *and constant* λ *satisfy* $\int_S m_*(\mathbf{x})\, d\mathbf{x} = 1$, *and minimize* (6).

Example 1. We consider an approximate accelerated failure model in survival analysis (Hosmer et al., 1998, p. 272). It is a generalized simple linear regression with $\mathbf{z}^{\mathrm{T}}(x) = (1, x)$, $\mu = \theta_0 + \theta_1 x$, and $h(\mu) = e^{\mu}$. By Theorem 5, the optimal design density has the form:

$$m_*(x; \theta_1) = \left[\frac{a_1x + a_2}{a_3x + a_4} + \frac{a_5}{e^{2\theta_1 x}(a_3x + a_4)^2}\right]^+. \tag{7}$$

Note that (7) is over-parameterized—if one of $a_1 - a_4$ is nonzero then we can assume that it is unity. The term $e^{2\theta_0}$ has been absorbed into a_5, but m_* still depends on θ_1. To address this issue we adopt a mixture of minimax and local approaches. We start at some $\theta_1 = \theta_1^{(0)}$. The corresponding optimal design density is $m_*^{(0)}(x)$. Then, we maximize (6) with $m = m_*^{(0)}$ over an interval containing $\theta_1^{(0)}$ to find the least favourable value of θ_1, say $\theta_1^{(1)}$. We iterate between minimizing over designs and maximizing over θ_1 until attaining convergence, say to θ_1^{LF}. Finally, we employ Theorem 5 to construct the "locally most robust" design density $m_*(x; \theta_1^{\mathrm{LF}})$.

To illustrate the approach, we consider the Class-H insulation data from Nelson (1990, Table 2.1). We transform the temperature variable t used there to our stress variable x with domain of [0, 1] via the transformation

$$x = \frac{-1.876 + 1000/(t + 273.15)}{0.283}.$$

The LS estimate for the nominal model is $\hat{\theta}_1 = 0.946$, with standard error 0.0486. A corresponding 99% confidence interval for θ_1 is (0.814, 1.08). Taking the model misspecification into account, we considered a broader region $\theta_1 \in [0, 2]$. We used the same extrapolation point $x_0 = 1.17$ as Nelson (1990). We carried out the process described above for several values of v, each time starting at $\theta_1^{(0)} = 0.946$. In each case we obtained $\theta_1^{\mathrm{LF}} = 2$. See Table 1 for the numerical values of the constants, and Fig. 1(a) for plots. As a comparison, Fig. 1(b) provides the plots of the locally optimal design densities at $\theta_1 = 0.946$. All plots use $a_4 = 1$ and $r_{T,S} = 1$.

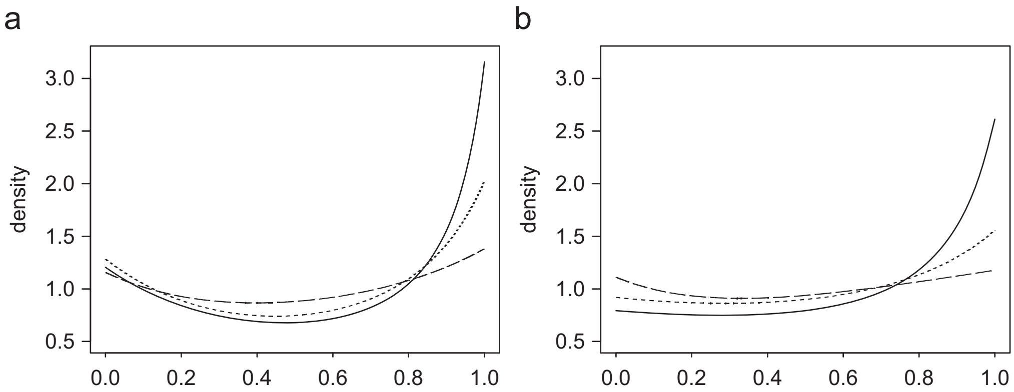

Fig. 1. Optimal minimax design densities $m_*(x;\theta_1)$ in Example 1 for $x_0 = 1.17$: (a) locally most robust design densities for $\theta_1 = \theta_1^{\mathrm{LF}}$ in [0, 2]; (b) locally most robust design densities for $\theta_1 = \theta_1^{(0)} = 0.946$. Each plot uses three values of $v : v = 2$ (solid line), $v = 1$ (dotted line), $v = 0.5$ (broken line).

5. Optimal designs with heteroscedasticity

Our problems P2 and P3 have become the following:

(P2) Find a density $m_*(\mathbf{x})$ which minimizes

$$\eta_S^{-2} \sup_{f\in\mathscr{F}, g\in\mathscr{G}} \mathrm{MSEE}(f, g, \mathbf{1}, m) = (\sqrt{\lambda_m} + r_{T,S})^2 + v\Omega^{-1/2}\left[\int_S \{l_m(\mathbf{x})m(\mathbf{x})\}^2\,\mathrm{d}\mathbf{x}\right]^{1/2} \tag{8}$$

with λ_m and $l_m(\mathbf{x})$ as defined in Theorems 1 and 3, respectively. Then $k_*(\mathbf{x}) = m_*(\mathbf{x})$ is the optimal one-point extrapolation design density for OLS estimation.

(P3) Find a density $m_*(\mathbf{x})$ which minimizes

$$\eta_S^{-2} \sup_{f\in\mathscr{F}, g\in\mathscr{G}} \mathrm{MSEE}(f, g, w_m, m) = (\sqrt{\lambda_m} + r_{T,S})^2 + v\Omega^{-1/2}\left[\int_S \{l_m(\mathbf{x})m^2(\mathbf{x})\}^{2/3}\,\mathrm{d}\mathbf{x}\right]^{3/2}. \tag{9}$$

Then the weights

$$w_*(\mathbf{x}) = \alpha_{m_*}\{l_{m_*}^2(\mathbf{x})m_*(\mathbf{x})\}^{-1/3} I[m_*(\mathbf{x}) > 0] \tag{10}$$

and the density

$$k_*(\mathbf{x}) = \alpha_{m_*}^{-1}[l_{m_*}(\mathbf{x})m_*^2(\mathbf{x})]^{2/3}, \tag{11}$$

with α_{m_*} defined in Theorem 3, are optimal for one-point extrapolation with WLS estimation.

5.1. Minimax designs for OLS: solution to P2

The solution to P2 is provided by Theorem 6 below.

Theorem 6. *The density $m_*(\mathbf{x})$ minimizing* (8) *for OLS estimation under heteroscedasticity is of the form*

$$m_*(\mathbf{x}) = [(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\gamma})(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}) + \lambda]^+ / [(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})^2\{1 + t(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})^2\}]. \tag{12}$$

The $p \times 1$ nonzero vectors $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, positive constant t, and constant λ satisfy $\int_S m_*(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$, and minimize (8).

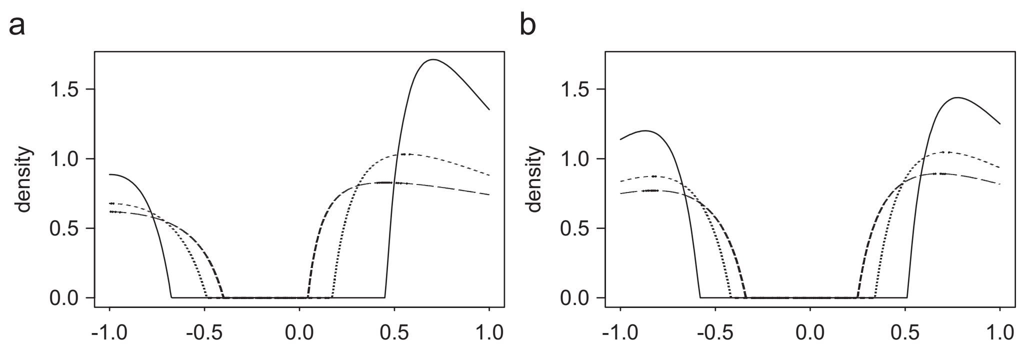

Fig. 2. Optimal minimax design densities $m_*(x)$ in Example 2 with $p = 1$: (a) $x_0 = 1.5$; (b) $x_0 = 5$. Each plot uses three values of v : $v = 10$ (solid line), $v = 1$ (dotted line), $v = 0.5$ (broken line).

Table 2
Numerical values for Example 2 with $p = 1$

x_0	v	a_1	a_2	a_3	a_5	a_6
1.5	0.25	3.78	0.853	5.96	−0.294	0.00116
	0.5	6.59	1.51	7.83	−2.43	0.00253
	1	14.06	3.18	11.23	−16.52	0.00320
	10	294.38	55.93	30.15	−2758.06	0.00537
	100	1247.25	223.60	39.53	−17826.34	0.00946
5	0.25	26.55	1.84	31.15	−34.92	0.000183
	0.5	54.39	3.72	45.58	−213.28	0.000188
	1	148.64	9.73	80.38	−525.93	0.000122
	10	751.39	42.48	66.96	−15013.05	0.00124
	100	2138.31	117.57	45.21	−32404.40	0.0150

Example 2. Consider an approximate polynomial regression model $E(Y|x) \approx \mu = \mathbf{z}^{\mathrm{T}}(x)\boldsymbol{\theta}_0 = \theta_0 + \theta_1 x + \cdots + \theta_p x^p$, where $\mathbf{z}^{\mathrm{T}}(x) = (1, x, \ldots, x^p)$ and the design space $S = [-1, 1]$. Applying Theorem 6 with $p = 1$ results in the form:

$$m_*(x) = \frac{[(a_1 x + a_2)(a_3 x + a_4) + a_5]^+}{(a_3 x + a_4)^2 + a_6(a_3 x + a_4)^4},$$

where $a_6 > 0$. Fig. 2 gives plots of the minimax extrapolation design densities for varying x_0 and v with $a_4 = 1$ when $r_{T,S} = 1$. A smaller v (more emphasis on bias) results in the minimax design becoming more uniform, while a larger v results in a design resembling that which minimizes variance alone. An extrapolation point x_0 closer to one end of the design space leads to more design points being placed on the corresponding side of the design space. As the distance between x_0 and S increases the design tends to become more symmetric. See Table 2 for some numerical values of the constants.

When $p = 2$, the minimax optimal design density has the form:

$$m_*(x) = \frac{[(a_0 + a_1 x + a_2 x^2)(b_0 + b_1 x + b_2 x^2) + c]^+}{(a_0 + a_1 x + a_2 x^2)^2 + d(a_0 + a_1 x + a_2 x^2)^4},$$

where $d > 0$. See Fig. 3 for plots of, and Table 3 for numerical values for, the minimax extrapolation design densities for varying $x_0 > 1$ and v with $a_0 = 1$ when $r_{T,S} = 1$. We observe the same qualitative features as when $p = 1$.

We now compare discrete implementations ξ_{*d} of ξ_* with two common competitors. Let ξ_{HL} be the Hoel–Levine design (Hoel and Levine, 1964) which was derived under the assumption of an exactly correct fitted model. When $p = 1$, the design points of ξ_{HL} are $x_1 = -1$ and $x_2 = +1$ with masses $\xi_{\mathrm{HL}}(-1) = (x_0 - 1)/2x_0$ and $\xi_{\mathrm{HL}}(1) = (x_0 + 1)/2x_0$. For $p = 2$, ξ_{HL} has three design points: $x_1 = -1$, $x_2 = 0$, and $x_3 = +1$ with masses $\xi_{\mathrm{HL}}(-1) = x_0(x_0 - 1)/2(2x_0^2 - 1)$,

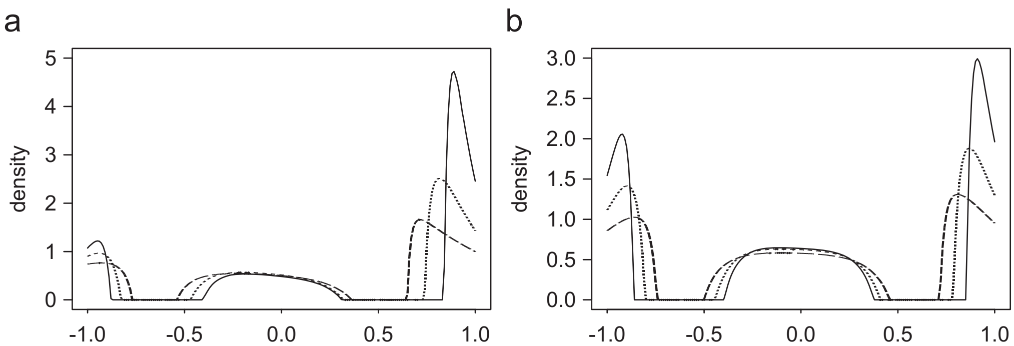

Fig. 3. Optimal minimax design densities $m_*(x)$ in Example 2 for $p=2$: (a) $x_0 = 1.5$; (b) $x_0 = 5$. Each plot uses three values of v : $v = 10$ (solid line), $v = 1$ (dotted line), $v = 0.25$ (broken line).

Table 3
Numerical values for Example 2 with $p = 2$

x_0	v	b_0	b_1	b_2	a_1	a_2	c	d
1.5	0.25	0.668	−0.521	−2.37	−0.192	−2.44	−0.102	0.123
	0.5	0.853	−0.644	−2.99	−0.127	−2.30	−0.200	0.280
	1	1.23	−0.858	−4.10	−0.0710	−2.17	−0.396	0.627
	10	8.20	−3.77	−24.49	0.0358	−1.92	−3.83	8.03
	100	77.71	−30.61	−223.96	0.0628	−1.86	−37.47	86.13
5	0.25	0.829	−0.144	−2.54	−0.0323	−2.29	−0.116	0.232
	0.5	1.12	−0.186	−3.40	−0.0191	−2.19	−0.237	0.472
	1	1.69	−0.255	−5.01	−0.0079	−2.11	−0.482	0.949
	10	11.68	−1.20	−31.87	0.0147	−1.92	−4.89	9.64
	100	111.06	−10.02	−296.67	0.0204	−1.88	−48.76	97.34

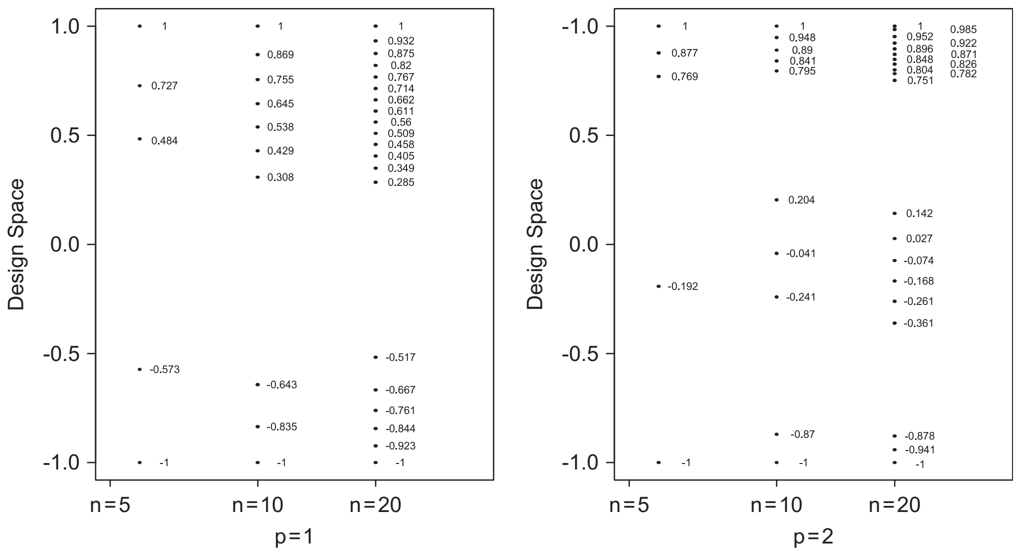

Fig. 4. Support points of the discretized designs $\xi_{*\mathrm{d}}$.

Table 4
Comparative values of b and v for three discrete designs: $\xi_{*\mathrm{d}}$, ξ_{Ud}, and ξ_{HL}

ν		b						v					
		$p=1$			$p=2$			$p=1$			$p=2$		
		$\xi_{*\mathrm{d}}$	ξ_{Ud}	ξ_{HL}	$\xi_{*\mathrm{d}}$	ξ_{Ud}	ξ_{HL}	$\xi_{*\mathrm{d}}$	ξ_{Ud}	ξ_{HL}	$\xi_{*\mathrm{d}}$	ξ_{Ud}	ξ_{HL}
1	$n=5$	0.909	1.049	1.275	2.086	2.145	2.284	7.179	10.224	4.531	32.025	41.45	22.891
	$n=10$	0.660	0.808	1.275	1.448	1.814	2.284	7.237	11.648	4.531	32.307	56.930	20.781
	$n=20$	0.477	0.596	1.275	1.054	1.414	2.284	7.441	12.417	4.510	33.609	66.885	.20.061
10	$n=5$	0.813	1.049	1.275	1.947	2.145	2.284	5.897	10.224	4.531	27.048	41.45	22.891
	$n=10$	0.592	0.808	1.275	1.450	1.814	2.284	6.024	11.648	4.531	29.070	56.930	20.781
	$n=20$	0.421	0.596	1.275	1.000	1.414	2.284	6.079	12.417	4.510	29.686	66.885	20.061

$\xi_{\mathrm{HL}}(0)=(x_0+1)(x_0-1)/(2x_0^2-1)$ and $\xi_{\mathrm{HL}}(1)=x_0(x_0+1)/2(2x_0^2-1)$. The frequencies $n\xi_{\mathrm{HL}}(\cdot)$ are then rounded to the nearest integer. Let ξ_{Ud} be the discrete uniform design on $[-1,1]$ with mass n^{-1} at each of the design points $-1+2(i-1)/(n-1)$ for $i=1,\ldots,n$. Our design $\xi_{*\mathrm{d}}$ places mass n^{-1} at each of $\xi_*^{-1}((i-1)/(n-1))$. We use $x_0=1.5$ and $\nu=1,10$ in the construction. For $n=5,10,20$, the support points of $\xi_{*\mathrm{d}}$ are shown in Fig. 4. In Table 4 we give some comparative values of

$$b=\|\mathbf{MZ}(\mathbf{Z}^{\mathrm{T}}\mathbf{MZ})^{-1}\mathbf{z}(x_0)\|,$$
$$v=\mathbf{z}^{\mathrm{T}}(x_0)(\mathbf{Z}^{\mathrm{T}}\mathbf{MZ})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{D}_g\mathbf{MZ}(\mathbf{Z}^{\mathrm{T}}\mathbf{MZ})^{-1}\mathbf{z}(x_0),$$

where $\mathbf{M}$ is the diagonal matrix of design masses, $\mathbf{Z}$ has rows $\mathbf{z}^{\mathrm{T}}(x_i)$ and $\mathbf{D}_g$ is the diagonal matrix with diagonal elements $g(x_i)$. Recall from Remark 2 that b is the component of the maximum, discretized EB which depends on the design; similarly, apart from the factor σ^2, v is the discretization of EV, from (3). We have rather arbitrarily chosen to evaluate v at $g(x)=1+x^2$. As one might expect, our design $\xi_{*\mathrm{d}}$ "wins" with respect to bias, while ξ_{HL} does so with respect to variance.

5.2. Minimax designs for WLS: solution to P3

The solution to P3 is provided by Theorem 7 below.

Theorem 7. *The minimizing $m_*(\mathbf{x})$ in (9) for WLS estimation is of the form*

$$m_*(\mathbf{x})=[B(\mathbf{x})-d(\mathbf{x})]^+/A^2(\mathbf{x}), \tag{13}$$

where

$$A(\mathbf{x})=\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta},\ B(\mathbf{x})=A(\mathbf{x})(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\gamma})+\lambda,$$

and where d satisfies the cubic equation

$$d^3+tA^2(\mathbf{x})d-tA^2(\mathbf{x})B(\mathbf{x})=0.$$

Explicitly,

$$d(\mathbf{x})=\left(\frac{t}{2}A^2(\mathbf{x})\right)^{1/3}[\{B(\mathbf{x})+\sqrt{C(\mathbf{x})}\}^{1/3}+\{B(\mathbf{x})-\sqrt{C(\mathbf{x})}\}^{1/3}],$$

with

$$C(\mathbf{x})=B^2(\mathbf{x})+\frac{4t}{27}A^2(\mathbf{x}).$$

The $p\times 1$ nonzero vectors $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and constants λ and $t>0$ satisfy $\int_S m_*(\mathbf{x})\,\mathrm{d}\mathbf{x}=1$, and minimize (9). *Then* (10) *and* (11) *provide the optimal one-point extrapolation design weights and design density for WLS estimation, respectively.*

Table 5
Numerical values for Example 3 with $p=1$, $r_{T,S}=1$, $S=[-1,1]$

x_0	v	a_1	a_2	a_3	t	λ
1.5	0.25	0.287	0.0136	0.388	0.0177	0.675
	0.5	−0.0118	0.00215	−0.0124	0.000161	0.541
	1	−0.255	1.70	−0.966	68.73	2.00
	10	8.69	8.78	53.96	8625.44	0.000230
	100	16.09	149.83	1023.40	47926350	−0.139
5	0.25	40.37	0.634	24.34	0.280	0.000618
	0.5	34.69	0.670	21.03	0.227	−0.00118
	1	18.99	1.30	21.97	31.95	−0.000154
	10	1.47	6.51	1.46	108419.4	−0.00368
	100	1.07	7.04	1.04	68740870	−0.367

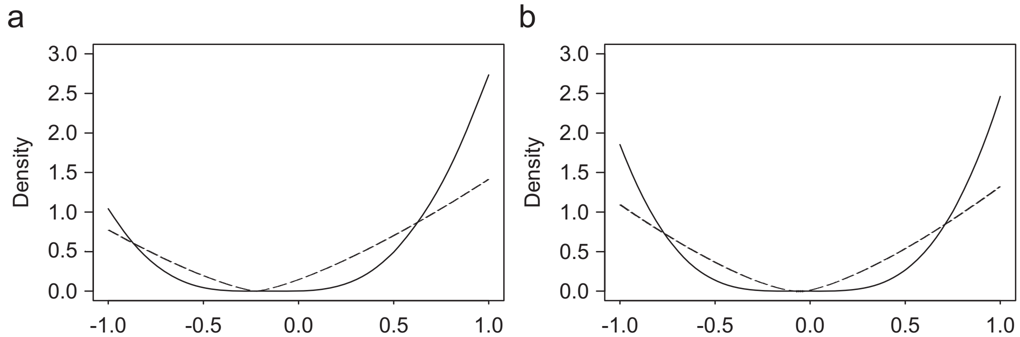

Fig. 5. Optimal extrapolation design densities for WLS and simple linear regression (Example 3): (a) $x_0=1.5$; (b) $x_0=5$. Each plot uses two values of v : $v=10$ (solid line), $v=0.5$ (dotted line).

Example 3. Consider an approximate polynomial model as in Example 2. By Theorem 7, the optimal minimax $m_*(x)=k_*(x)w_*(x)$ for WLS has the form (13) with

$$A(x)=\sum_{i=0}^{p}\beta_i x^i, \quad B(\mathbf{x})=A(x)\sum_{i=0}^{p}\gamma_i x^i+\lambda.$$

The minimax design ξ_* has density $k_*(x)$ computed from (11). The minimax weights $w_*(x)$ are obtained from (10). Assuming a nonzero intercept we can without loss of generality take $\beta_0=1$.

For $p=1$ we have $A(x)=1+a_1x$ and $B(x)=A(x)(a_2+a_3x)+\lambda$. See Table 5 for numerical values of the constants. Fig. 5 gives plots of the minimax extrapolation design densities for $S=[-1,1]$ and varying $x_0>1$. For $p=2$ we have $A(x)=1+a_1x+a_2x^2$ and $B(x)=A(x)(a_3+a_4x+a_5x^2)+\lambda$. See Table 6 for some numerical values of the constants, and Fig. 6 for plots.

6. Optimal unbiased designs: solution to P4

We say that a design/weights pair (ξ, w) is *unbiased* if it satisfies $\mathrm{EB}(f,w,\xi)=f^2(\mathbf{x}_0)$ for all $f\in\mathscr{F}$, so that $\sup_{f\in\mathscr{F}}\mathrm{EB}(f,w,\xi)=\eta_T^2$. The following theorem, which is essentially Theorem 2.2(b) of Fang and Wiens (1999), gives a necessary and sufficient condition for unbiasedness.

Theorem 8. *The pair* (w,ξ) *is unbiased if and only if* $m(\mathbf{x})\equiv\Omega=1/\int_S \mathrm{d}\mathbf{x}$.

Table 6
Numerical values for Example 3 with $p = 2$, $r_{T,S} = 1$, $S = [-1, 1]$

x_0	v	a_1	a_2	a_3	a_4	a_5	t	λ
1.5	0.25	0.469	2.47	0.00250	0.411	2.18	0.0300	0.767
	0.5	0.514	2.83	0.199	0.751	4.18	5.02	1.73
	1	0.533	3.15	0.0207	1.64	9.80	141.91	4.87
	10	−1.25	0.466	0.000150	−1.52	−0.30	0.0000641	0.477
	100	−2.96	1.92	0.888	−1.90	0.617	0.034	−0.00813
5	0.25	0.0665	1.76	0.173	0.122	3.38	10.87	2.19
	0.5	0.0779	2.14	0.288	0.376	10.97	645.82	7.55
	1	0.0878	2.66	0.962	12.87	416.62	36,493,560	285.96
	10	−0.000741	−0.603	0.326	0.210	−11.52	759.74	7.84
	100	−0.00100	−0.617	0.220	0.632	−36.26	21033.71	23.69

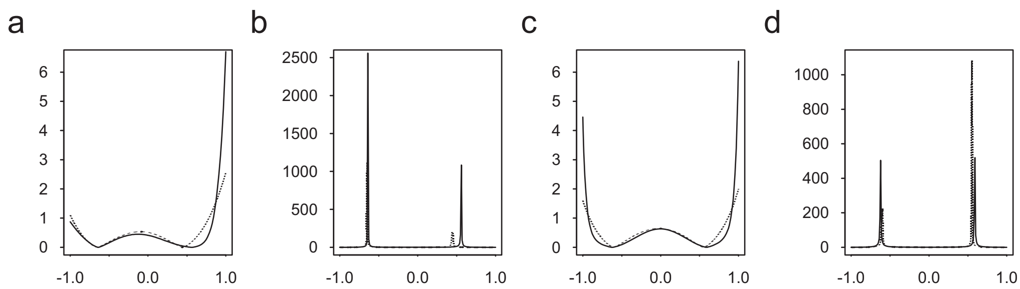

Fig. 6. Optimal extrapolation design densities and minimax weights for WLS and quadratic regression (Example 3): (a) design densities for $x_0 = 1.5$; (b) minimax weights for $x_0 = 1.5$; (c) design densities for $x_0 = 5$; (d) minimax weights for $x_0 = 5$. Each plot uses two values of v : $v = 10$ (solid line), $v = 0.1$ (dotted line).

We can construct the optimal unbiased extrapolation design $m_0(\mathbf{x})$ by forcing $\sup_{f\in\mathscr{F}} \mathrm{EB}(f, w, \xi) = \eta_T^2$ through the choice $k = \Omega/w$, and then minimizing $\sup_{g\in\mathscr{G}} \mathrm{EV}(g, w, \xi)$ over w. From Theorem 4, the optimal weight function is

$$w_0(\mathbf{x}) = \Omega\alpha_0[\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)]^{-4/3},$$

and the optimal unbiased extrapolation design density is

$$k_0(\mathbf{x}) = \alpha_0^{-1}[\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)]^{4/3},$$

with

$$\alpha_0 = \int_S [\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)]^{4/3}\,\mathrm{d}\mathbf{x}.$$

The minimax MSEE is

$$\min_{(w,\xi)} \sup_{f\in\mathscr{F}, g\in\mathscr{G}} \mathrm{MSEE}(f, g, w, m) = \eta_T^2 + \min_{(w,\xi)} \sup_{g\in\mathscr{G}} \mathrm{EV}(g, w_m, \xi) = \eta_S^2\{r_{T,S}^2 + v\Omega^{-1/2}\alpha_0^{3/2}\}.$$

We summarize these observations below.

Theorem 9. *The density $k_0(\mathbf{x})$ of the optimal unbiased one-point extrapolation design measure ξ_0, and optimal weights w_0, which minimize $\sup_{f\in\mathscr{F}, g\in\mathscr{G}}\mathrm{MSEE}(f, g, w, \xi)$ subject to $\sup_{f\in\mathscr{F}}\mathrm{EB}(f, w, \xi) = \eta_T^2$ are given by*

$$k_0(\mathbf{x}) = \frac{[\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)]^{4/3}}{\int_S[\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)]^{4/3}\mathrm{d}\mathbf{x}},$$

and $w_0(\mathbf{x}) = \Omega/k_0(\mathbf{x})$. *Minimax* MSEE *is*

$$\sup_{f\in\mathscr{F},g\in\mathscr{G}} \mathrm{MSEE}(f, g, w_0, \xi_0) = \eta_S^2 \left\{ r_{T,S}^2 + v\Omega^{-1/2}\left[\int_S [\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)]^{4/3}\,\mathrm{d}\mathbf{x}\right]^{3/2}\right\}, \tag{14}$$

attained at $g_0(\mathbf{x}) = w_0^{-1/2}(\mathbf{x})$.

Example 4. Consider an approximate log-linear multiple regression model $E(Y|\mathbf{x}) \approx \exp(\mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\theta}_0) = \exp(\theta_0 + \theta_1 x_1 + \cdots + \theta_q x_q)$.

Note that the designs provided by Theorem 9 for this example depend on $\boldsymbol{\theta}_1 = (\theta_1, \ldots, \theta_q)^{\mathrm{T}}$ but not on θ_0. As in Example 1 we can find locally most robust designs in a neighbourhood Θ of a starting value $\boldsymbol{\theta}_1^{(0)}$. We first construct the design $k_0(\mathbf{x}, \boldsymbol{\theta}_1^{(0)})$ and weights $\Omega/k_0(\mathbf{x}, \boldsymbol{\theta}_1^{(0)})$ provided by Theorem 9. We then find the least favourable $\boldsymbol{\theta}_{1,\mathrm{LF}}$ in Θ. From Theorem 3, we find that this is equivalent to maximizing

$$\int_S \frac{\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{A}_S^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0)\}^4}{k_0^2(\mathbf{x}, \boldsymbol{\theta}^{(0)})}\,\mathrm{d}\mathbf{x}$$

over the occurrences of $\boldsymbol{\theta}_1$ in the numerator of the integrand. We then construct the unbiased optimal design for $\boldsymbol{\theta}_{\mathrm{LF}}$, and iterate to convergence.

When $q = 1$, $S = [-0.5, 0.5]$ the unbiased minimax design density is

$$k_0(x) = \begin{cases} \dfrac{\{e^{\theta_{\mathrm{LF}}x}[(c - bx_0) + (ax_0 - b)x]\}^{4/3}}{\int_{-0.5}^{0.5}\{e^{\theta_{\mathrm{LF}}x}[(c - bx_0) + (ax_0 - b)x]\}^{4/3}\,\mathrm{d}x}, & \text{when } \theta_{\mathrm{LF}} \neq 0, \\ \dfrac{(1 + 12x_0x)^{4/3}}{\int_{-0.5}^{0.5}(1 + 12x_0x)^{4/3}\,\mathrm{d}x}, & \text{when } \theta_{\mathrm{LF}} = 0, \end{cases}$$

where

$$a = 2\sinh(\theta_{\mathrm{LF}}), \quad b = \cosh(\theta_{\mathrm{LF}}) - a, \quad c = 0.5\sinh(\theta_{\mathrm{LF}}) - 2b.$$

For a simple demonstration of the procedure described above, we take $\Theta = [0.5, 0.7]$ and consider the cases $x_0 = \pm 2$ and $x_0 = \pm 9$. For both $x_0 = 2$ and $x_0 = 9$, the iterates converge to $\theta_{\mathrm{LF}} = 0.7$. The unbiased minimax design density at $x_0 = 2$ is

$$k_0(x) = 0.447\{e^{0.7x}(1.427 + 3.296x)\}^{4/3},$$

and that at $x_0 = 9$ is

$$k_0(x) = 0.104\{e^{0.7x}(3.261 + 13.917x)\}^{4/3}.$$

When $x_0 = -2$ and -9, we find $\theta_{\mathrm{LF}} = 0.5$. The unbiased minimax design density at $x_0 = -2$ is

$$k_0(x) = 2.063\{e^{0.5x}(0.261 - 2.170x)\}^{4/3},$$

and that at $x_0 = -9$ is

$$k_0(x) = 0.293\{e^{0.5x}(0.859 - 9.465x)\}^{4/3}.$$

The corresponding optimal weights are $w_0(x) = 1/k_0(x)$.

Example 5. Consider an approximate polynomial regression model $E(Y|x) \approx \mathbf{z}^{\mathrm{T}}(x)\boldsymbol{\theta}_0 = \theta_0 + \theta_1 x + \cdots + \theta_p x^p$ with $S = [-1, 1]$. By Theorem 9, the unbiased optimal density is

$$k_0(x) = \frac{[\mathbf{z}^{\mathrm{T}}(x)A_S^{-1}\mathbf{z}(x_0)]^{4/3}}{\int_S [\mathbf{z}^{\mathrm{T}}(x)A_S^{-1}\mathbf{z}(x_0)]^{4/3}\,\mathrm{d}x}$$

with optimal weights $w_0(x) = 0.5/k_0(x)$.

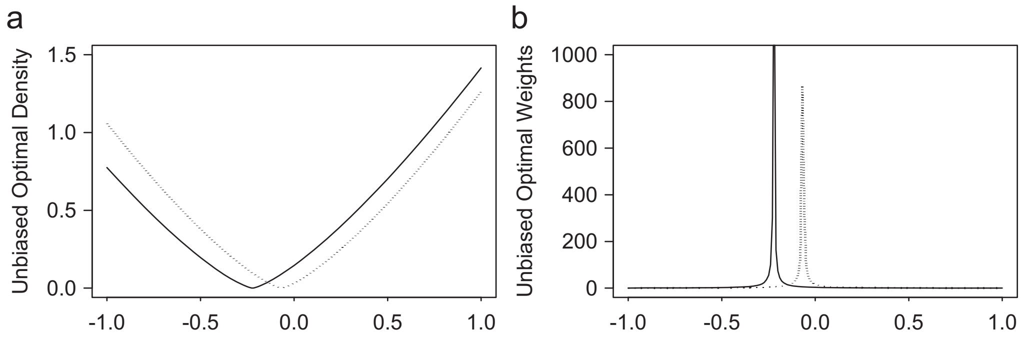

Fig. 7. Unbiased optimal densities and weights for SLR (Example 5): (a) design densities; (b) weights. Each plot uses two values of x_0 : $x_0 = 1.5$ (solid line) and $x_0 = 5$ (dotted line).

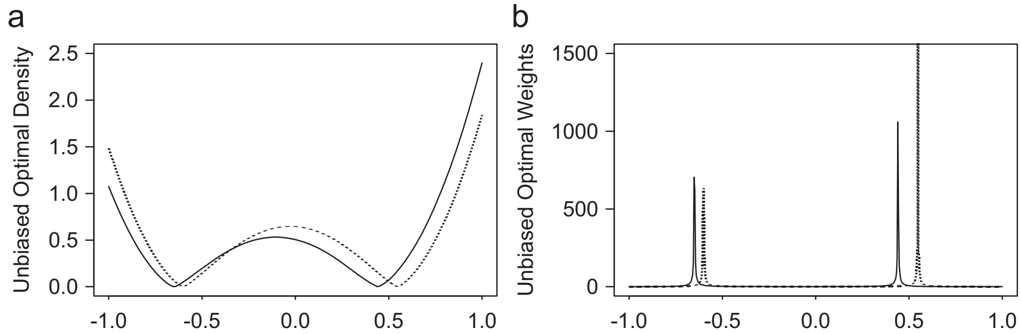

Fig. 8. Unbiased optimal design densities and weights for quadratic regression (Example 5): (a) densities; (b) weights. Each plot uses two values of x_0: $x_0 = 1.5$ (solid line) and $x_0 = 10$ (dotted line).

When $p = 1$, we have the design

$$k_0(x) = \frac{3.5x_0(0.5 + 1.5x_0x)^{4/3}}{(0.5 + 1.5x_0)^{7/3} - (0.5 - 1.5x_0)^{7/3}}.$$

When $p = 2$, the design is

$$k_0(x) \propto \{(3 - 5x_0^2) + 4x_0x + (15x_0^2 - 5)x^2\}^{4/3}.$$

See Figs. 7 and 8.

7. Comparisons and remarks

In Examples 2 and 3, we compared our designs for P2 and P3 with two more conventional competing designs ξ_{HL} and ξ_{U}. In this section, we use the approximate polynomial models ($p = 1, 2$) of these examples to compare the robust minimax designs for P2–P4 with each other and again with ξ_{HL} and the continuous uniform design ξ_{U}. Let $\xi^{(2)}$, $\xi^{(3)}$ and $\xi^{(4)}$ denote the robust optimal designs that we obtained for P2, P3 and P4, respectively. Table 7 gives the comparative values of $\eta_S^{-2}\mathrm{EV}$ when there is no contamination and Table 8 gives those of $\eta_S^{-2}\sup_{f,g}\mathrm{MSEE}$ for ξ_{U} when there is maximal contamination. Of course, $\sup_{f,g}\mathrm{MSEE}$ for ξ_{HL} is infinite.

When there is no contamination, we denote by $\mathrm{re}_{\mathrm{HL}}^{(0)}(\xi^{(\cdot)})$ the efficiencies of $\xi^{(2)}$, $\xi^{(3)}$ and $\xi^{(4)}$ relative to ξ_{HL} and by $\mathrm{re}_{\mathrm{U}}^{(0)}(\xi^{(\cdot)})$ the efficiencies relative to ξ_{U}. Under maximal contamination we write instead $\mathrm{re}_{\mathrm{U}}^{(\max)}(\xi^{(\cdot)})$ and $\mathrm{re}_{\mathrm{HL}}^{(\max)}(\xi^{(\cdot)}) = \infty$. Table 9 provides the relative efficiencies $\mathrm{re}_{\mathrm{HL}}^{(0)}$ and $\mathrm{re}_{\mathrm{U}}^{(0)}$ while Table 10 provides the relative efficiencies $\mathrm{re}_{\mathrm{U}}^{(\max)}(\xi^{(\cdot)})$.

Table 7
Comparative values of η_S^{-2}EV when there is no contamination of the polynomial regression model (Examples 2 and 3)

x_0	v	$p=1$			$p=2$			$\xi_{\mathrm{HL}};\ p=1(2)$	$\xi_U;\ p=1(2)$
		$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$		
1.5	0.25	1.57	1.44	1.44	7.30	7.96	8.41	0.563(3.06)	1.94(12.27)
	0.5	2.70	2.88	2.88	13.16	15.29	16.82	1.13(6.13)	3.88(24.54)
	1	4.71	5.68	5.76	23.87	28.82	33.65	2.25(12.25)	7.75(49.08)
	10	36.47	42.00	57.62	198.74	261.10	336.49	22.5(122.5)	77.5(490.78)
	100	347.05	398.60	576.19	1923.71	2424.7	3364.90	225(1225)	775(4907.81)
5	0.25	13.77	14.14	14.60	1140.24	1240.23	1311.03	6.25(600.25)	19(1730.25)
	0.5	24.81	28.27	29.21	2112.96	2387.43	2622.06	12.5(1200.5)	38(3460.5)
	1	45.28	51.06	58.41	3950.39	4564.29	5244.11	25(2401.00)	76(6921)
	10	383.80	432.99	584.13	35092.10	42875.19	52441.14	250(24010)	760(69210)
	100	3726.18	4197.94	5841.26	343816.60	417639.3	524411.4	2500(240100)	7600(692100)

Table 8
Comparative values of $\eta_S^{-2}\sup_{f,g}$MSEE when there is maximal contamination of the polynomial regression model (Examples 2 and 3)

x_0	v	$p=1$			$p=2$			$\xi_U;\ p=1(2)$
		$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	
1.5	0.25	3.50	2.59	2.59	19.50	14.71	10.39	3.78(23.10)
	0.5	5.58	4.17	4.17	30.85	23.61	19.79	6.56(45.20)
	1	9.27	7.27	7.35	51.63	40.85	38.58	12.12(89.40)
	10	66.91	58.24	64.49	384.07	336.75	376.77	112.15(884.98)
	100	629.58	552.78	635.90	3629.60	3211.04	3758.73	1112.55(8840.79)
5	0.25	22.42	16.65	16.95	1833.70	1422.04	1430.95	26.71(2586.07)
	0.5	40.69	32.30	32.90	3410.07	2780.92	2860.90	52.43(5171.13)
	1	74.91	62.85	64.80	6405.83	5432.51	5720.81	103.86(10341.27)
	10	648.81	575.17	638.96	57275.39	52036.29	57199.07	1029.60(103403.65)
	100	6330.03	5643.09	6380.57	561056.9	512051.6	571981.67	10286.98(1034027.5)

We have provided methods of constructing optimally robust designs for one-point regression extrapolation, taking into account various model uncertainties. The results require extensive numerical work prior to implementation. However, we can give some informal guidelines:

1. As v increases, the designs place more emphasis on variance minimization and less on protection from bias. As we would expect, the experimenter should then place relatively more design points closer to the boundary of the design space. With respect to the position of the extrapolation point relative to the design space, the experimenter should place relatively more design points in that segment of the design space which is closer to the extrapolation point, with this prescription becoming more emphatic when the extrapolation point is close to the design space.
2. Compared to designs for variance minimization alone, the designs we have found in this work are substantially more uniform. They can roughly be described as being obtained by replacing the point masses of the variance minimizing designs by uniform densities on regions containing, but not restricted to, these atoms.
3. Under heteroscedasticity the designs for P3 are, as expected, the most efficient. The gains in efficiency are greater when v is at least moderately large. Particularly for small v, the numerical simplicity of the designs for P4 makes them attractive competitors.

Table 9
Relative efficiencies $\mathrm{re}_{\mathrm{HL}}^{(0)}(\xi^{(\cdot)})$ and $\mathrm{re}_{\mathrm{U}}^{(0)}(\xi^{(\cdot)})$ for the polynomial regression model (Examples 2 and 3)

x_0	v	$p=1$			$p=2$		
		$\xi^{(2)}$: HL/U	$\xi^{(3)}$: HL/U	$\xi^{(4)}$: HL/U	$\xi^{(2)}$: HL/U	$\xi^{(3)}$: HL/U	$\xi^{(4)}$: HL/U
1.5	0.25	0.36/1.14	0.39/1.35	0.39/1.35	0.42/1.68	0.38/1.54	0.36/1.46
	0.5	0.42/1.44	0.39/1.35	0.39/1.35	0.47/1.86	0.40/1.60	0.36/1.46
	1	0.48/1.65	0.40/1.36	0.39/1.35	0.51/2.06	0.43/1.70	0.36/1.46
	10	0.62/2.13	0.54/1.85	0.39/1.35	0.62/2.47	0.47/1.88	0.36/1.46
	100	0.65/2.23	0.56/1.94	0.39/1.35	0.64/2.55	0.51/2.02	0.36/1.46
5	0.25	0.45/1.38	0.44/1.34	0.43/1.30	0.53/1.52	0.48/1.40	0.46/1.32
	0.5	0.50/1.53	0.44/1.34	0.43/1.30	0.57/1.64	0.50/1.45	0.46/1.32
	1	0.55/1.68	0.49/1.49	0.43/1.30	0.61/1.75	0.53/1.52	0.46/1.32
	10	0.65/1.98	0.58/1.76	0.43/1.30	0.68/1.97	0.56/1.61	0.46/1.32
	100	0.67/2.04	0.59/1.81	0.43/1.30	0.70/2.01	0.57/1.66	0.46/1.32

Table 10
Relative efficiencies $\mathrm{re}_{\mathrm{U}}^{(\max)}(\xi^{(\cdot)})$ for the polynomial regression model (Examples 2 and 3)

x_0	v	$p=1$			$p=2$		
		$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$
1.5	0.25	1.08	1.46	1.46	1.18	1.57	2.22
	0.5	1.18	1.57	1.57	1.47	1.91	2.28
	1	1.31	1.67	1.65	1.73	2.19	2.32
	10	1.68	1.93	1.74	2.30	2.63	2.35
	100	1.77	2.01	1.75	2.44	2.75	2.35
5	0.25	1.19	1.60	1.58	1.41	1.82	1.81
	0.5	1.29	1.62	1.59	1.52	1.86	1.81
	1	1.39	1.65	1.60	1.61	1.90	1.81
	10	1.59	1.79	1.61	1.81	1.99	1.81
	100	1.63	1.82	1.61	1.84	2.02	1.81

Acknowledgements

The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada. The authors would like to thank two anonymous referees for their constructive criticism on the previous versions of this paper.

Appendix. Derivations

The proof of Theorem 5 is very similar to but simpler than that of Theorem 6, and so is omitted.

Proof of Theorem 6. We seek a nonnegative function $m(\mathbf{x})$ minimizing (8) subject to $\int_S m(\mathbf{x})\,\mathrm{d}\mathbf{x}=1$. For a Lagrange multiplier s it is necessary and sufficient that m minimize

$$(\sqrt{\lambda_m}+r_{T,S})^2+v\Omega^{-1/2}\left[\int_S\{l_m(\mathbf{x})m(\mathbf{x})\}^2\,\mathrm{d}\mathbf{x}\right]^{1/2}-2s\int_S m(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

among all densities, and satisfy the side condition. After a lengthy calculation we obtain the first-order condition

$$I:=\int_S\{P(\mathbf{x})m(\mathbf{x})-Q(\mathbf{x})-u\}(m(\mathbf{x})-m_1(\mathbf{x}))\,\mathrm{d}\mathbf{x}\geqslant 0 \tag{A.1}$$

for all densities m_1, where

$$P(\mathbf{x}) = (\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})^2[1 + t(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})^2] \quad \text{and} \quad Q(\mathbf{x}) = (\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\gamma)(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}),$$

for

$$\begin{aligned}
\boldsymbol{\beta} &= \mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}_0),\\
t &= \frac{v}{2}\Omega^{-1/2}\left(1 + \frac{r_{T,S}}{\sqrt{\lambda_m}}\right)^{-1}\left[\int_S \{l_m(\mathbf{x})m(\mathbf{x})\}^2\,\mathrm{d}\mathbf{x}\right]^{-1/2} > 0,\\
\gamma &= \mathbf{B}^{-1}\left\{\mathbf{K} + \frac{t}{2}\left[\int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})l_m(\mathbf{x})m^2(\mathbf{x})\,\mathrm{d}\mathbf{x}\right]\right\}\boldsymbol{\beta},\\
u &= \left(1 + \frac{r_{T,S}}{\sqrt{\lambda_m}}\right)^{-1} s.
\end{aligned}$$

To see the consequences of (A.1), write S^+ for the subset of S on which $m(\mathbf{x}) > 0$, and $S^0 = S \backslash S^+$. Let $c = \sup_S\{P(\mathbf{x})m(\mathbf{x}) - Q(\mathbf{x}) - u\}$, let $\{\mathbf{x}_j\}$ be a sequence of points in S^+ with $P(\mathbf{x}_j)m(\mathbf{x}_j) - Q(\mathbf{x}_j) - u$ approaching c, and consider a sequence $\{m_{1j}\}$ of point masses at $\mathbf{x}_j$. Then for this sequence (A.1) implies

$$\int_{S^+}\{P(\mathbf{x})m(\mathbf{x}) - Q(\mathbf{x}) - u\}m(\mathbf{x})\,\mathrm{d}\mathbf{x} \geqslant c \geqslant \sup_{S^+}\{P(\mathbf{x})m(\mathbf{x}) - Q(\mathbf{x}) - u\},$$

so that in particular $P(\mathbf{x})m(\mathbf{x}) - Q(\mathbf{x}) - u \equiv c$ on S^+ and $-Q(\mathbf{x}) - u = P(\mathbf{x})m(\mathbf{x}) - Q(\mathbf{x}) - u \leqslant c$ on S^0. Thus

$$m(\mathbf{x}) = \frac{Q(\mathbf{x}) + u + c}{P(\mathbf{x})}, \quad \mathbf{x} \in S^+. \tag{A.2}$$

Conversely, if (A.2) holds and $Q(\mathbf{x}) + u + c \geqslant 0$ on S^0 then

$$\begin{aligned}
I &= c\int_{S^+}(m(\mathbf{x}) - m_1(\mathbf{x}))\,\mathrm{d}\mathbf{x} + \int_{S^0}[Q(\mathbf{x}) + u]m_1(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
&= c - c\int_{S^+} m_1(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{S^0}[Q(\mathbf{x}) + u + c]m_1(\mathbf{x})\,\mathrm{d}\mathbf{x} - c\int_{S^0} m_1(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
&= c - c\int_S m_1(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{S^0}[Q(\mathbf{x}) + u + c]m_1(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
&\geqslant 0,
\end{aligned}$$

satisfying (A.1). Thus, in order that (A.1) hold, it is necessary and sufficient that (A.2) hold, for any c such that the right-hand side of (A.2) is nonnegative throughout S. More generally, m has the form:

$$m(\mathbf{x}) = \frac{[Q(\mathbf{x}) + u + c]^+}{P(\mathbf{x})} = \frac{[(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\gamma)(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}) + \lambda]^+}{(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})^2[1 + t(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})^2]}, \tag{A.3}$$

with $\lambda = u + c$. Of course, $\boldsymbol{\beta}$, γ, λ and t themselves depend on m. Rather than solve (A.3) for m it is simpler merely to choose these constants so as to satisfy $\int_S m(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$ and minimize (8). □

Proof of Theorem 7. As in the preceding proof, we seek a density $m(\mathbf{x})$ minimizing

$$(\sqrt{\lambda_m} + r_{T,S})^2 + v\Omega^{-1/2}\left\{\int_S [l_m(\mathbf{x})m^2(\mathbf{x})]^{2/3}\,\mathrm{d}\mathbf{x}\right\}^{3/2} - 2s\int_S m(\mathbf{x})\,\mathrm{d}\mathbf{x},$$

this leads to

$$I := \int_S\{a(\mathbf{x})m^{1/3}(\mathbf{x}) + b(\mathbf{x})m(\mathbf{x}) - c(\mathbf{x}) - u\}(m - m_1)\,\mathrm{d}\mathbf{x} \geqslant 0,$$

where u and $\boldsymbol{\beta}$ are as before and

$$a(\mathbf{x}) = t^{1/3}(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})^{4/3} > 0,$$
$$b(\mathbf{x}) = l_m(\mathbf{x}) = (\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})^2 > 0,$$
$$c(\mathbf{x}) = (\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\gamma),$$

for

$$t = v^3\Omega^{-3/2}\left\{1 + \frac{r_{T,S}}{\sqrt{\lambda_m}}\right\}^{-3}\left[\int_S l_m^{2/3}(\mathbf{x})m^{4/3}(\mathbf{x})\,\mathrm{d}\mathbf{x}\right]^{3/2} > 0,$$
$$\gamma = \mathbf{B}^{-1}\left\{\mathbf{K} + t^{1/3}\int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})l_m^{-1/3}(\mathbf{x})m^{4/3}(\mathbf{x})\,\mathrm{d}\mathbf{x}\right\}\boldsymbol{\beta}.$$

Continuing as in the proof of Theorem 6 we find that on S^+,

$$a(\mathbf{x})m^{1/3}(\mathbf{x}) + b(\mathbf{x})m(\mathbf{x}) - c(\mathbf{x}) - \lambda \equiv 0 \tag{A.4}$$

for a constant λ. To solve (A.4) let $d(\mathbf{x}) = c(\mathbf{x}) + \lambda - b(\mathbf{x})m$, obtaining $d^3 + (a^3/b)d - a^3(c+\lambda)/b = 0$. Since the linear coefficient is positive there is only one real root, which can be obtained from Cardano's formula (Dunham, 1990). The proof is completed upon writing $d(\mathbf{x})$ in terms of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, λ and t. □

References

Dette, H., Wong, W.-K., 1996. Robust optimal extrapolation designs. Biometrika 83, 667–680.

Draper, N.R., Herzberg, A., 1973. Some designs for extrapolation outside a sphere. J. Roy. Statist. Soc. Ser. B 35, 268–276.

Dunham, W., 1990. Cardano and the solution of the cubic. Journey through Genius: The Great Theorems of Mathematics. Wiley, New York, pp. 135–154. (Chapter 6).

Fang, Z., Wiens, D.P., 1999. Robust extrapolation designs and weights for biased regression models with heteroscedastic errors. Canad. J. Statist. 27, 751–770.

Fang, Z., Wiens, D.P., 2000. Integer-valued minimax robust designs for estimation and extrapolation in heteroscedastic approximately linear models. J. Amer. Statist. Assoc. 95, 807–818.

Ford, I., Kitsos, C.P., Titterington, D.M., 1989. Recent advances in nonlinear experimental design. Technometrics 31, 49–60.

Heo, G., Schmuland, B., Wiens, D.P., 2001. Restricted minimax robust designs for misspecified regression models. Canad. J. Statist. 29, 117–128.

Hoel, P.G., Levine, A., 1964. Optimal spacing and weighting in polynomial prediction. Ann. Math. Statist. 35, 1553–1560.

Hosmer Jr., D.W., Lemeshow, S., 1998. Applied Survival Analysis. Wiley, New York.

Huang, M.N.L., Studden, W.J., 1988. Model robust extrapolation designs. J. Statist. Plann. Inference 18, 1–24.

Huber, P.J., 1975. Robustness and designs. In: Srivastava, J.N. (Ed.), A Survey of Statistical Design and Linear Models. North Holland, Amsterdam, pp. 287–303.

Nelson, W., 1990. Accelerated Testing: Statistical Models, Test Plans, and Data Analyses. Wiley, New York.

Seber, G.A.F., Wild, C.J., 2003. Nonlinear Regression. Wiley, New York.

Sinha, S., Wiens, D.P., 2002. Robust sequential designs for nonlinear regression. Canad. J. Statist. 30, 601–618.

Spruill, M.C., 1984. Optimal designs for minimax extrapolation. J. Multivariate Anal. 15, 52–62.

Wiens, D.P., 1992. Minimax designs for approximately linear regression. J. Statist. Plann. Inference 31, 353–371.

Yue, R.-X., Hickernell, F.J., 1999. Robust designs for fitting linear models with misspecification. Statist. Sin. 9, 1053–1069.