Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 30–46
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Robust prediction and extrapolation designs for misspecified generalized linear regression models

Douglas P. Wiens*, Xiaojian Xu

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alta.,Canada T6G 2G1

Available online 23 May 2007

Abstract

We study minimax robust designs for response prediction and extrapolation in biased linear regression models. We extend previous work of others by considering a nonlinear fitted regression response, by taking a rather general extrapolation space and, most significantly, by dropping all restrictions on the structure of the regressors. Several examples are discussed.

MSC: primary 62K0562F35; Secondary 62J12

Keywords: Regression design; Nonlinear least squares; Generalized linear response; Heteroscedasticity; Nonsmooth optimization

1. Introduction

In this article, we investigate the construction of robust designs for both prediction and extrapolation of regression responses. In our framework the response fitted by the experimenter is a known 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 regression response, and for possible heteroscedasticity.

Consider a regression model

$$E(Y|\mathbf{x}) \approx h(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x})), \tag{1}$$

for a q-dimensional vector $\mathbf{x}$ belonging to a bounded design space S and for p regressors $\mathbf{z}(\mathbf{x})=(z_1(\mathbf{x}), z_2(\mathbf{x}), \ldots, z_p(\mathbf{x}))^{\mathrm{T}}$. The function h is strictly monotonic with a bounded second derivative. We assume that $\|\mathbf{z}(\mathbf{x})\|$ is bounded on S. As indicated in (1), the fitted response is typically acknowledged to be only an approximation. The least squares estimates $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ and $\hat{Y} = h(\mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\theta})$ of $E(Y|\mathbf{x})$ are possibly biased if the response is misspecified. In this situation, robust designs can play an important role in choosing optimal design points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in S$ so that estimates $\hat{\boldsymbol{\theta}}$ and $\hat{Y}$ remain relatively efficient, with small bias caused by the model misspecification.

The true model may be written

$$E(Y|\mathbf{x}) = h(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x})) + n^{-1/2}f(\mathbf{x}), \tag{2}$$

* Corresponding author. Tel.: +1 780 4924406; fax: +1 780 4926826.
E-mail addresses: doug.wiens@ualberta.ca (D.P. Wiens), xiaojian@ualberta.ca (X. Xu).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2007.05.025

where the contaminant f is unknown but 'small'. This may be viewed as arising from imprecision in the specification of h, or it can arise from a misspecified linear term and a two-term Taylor expansion: $h(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x})+\phi(\mathbf{x})) \approx h(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}))+h'(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}))\phi(\mathbf{x})=h(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}))+n^{-1/2}f(\mathbf{x})$. The factor $n^{-1/2}$ is necessary for an appropriate asymptotic treatment—see Wiens and Xu (2005).

The experimenter takes n uncorrelated observations $Y_i = Y(\mathbf{x}_i)$, with $\mathbf{x}_i$ freely chosen from a design space S. One possible goal is prediction, or equivalently the estimation of $E(Y|\mathbf{x})$ throughout the region $T=S$. If instead $T \cap S = \phi$, the goal is extrapolation. In this article, we discuss both prediction problems and extrapolation problems. We will as well allow for the possibility that observations on Y, although uncorrelated, are heteroscedastic: $var\{Y(\mathbf{x})\} = \sigma^2 g(\mathbf{x})$ for an unknown function within a certain class. We estimate $\boldsymbol{\theta}$ by nonlinear least squares, possibly weighted with nonnegative weights $w(\mathbf{x})$.

For the prediction case, our loss function is n times the integrated mean squared prediction error (IMSPE) of $\hat{Y}(\mathbf{x})$ in estimating $E(Y|\mathbf{x})$, $\mathbf{x} \in S$. For extrapolation, loss is n times the integrated mean squared extrapolation error (IMSEE) of $\hat{Y}(\mathbf{x})$ in estimating $E(Y|\mathbf{x})$, $\mathbf{x} \in T$. Both depend on the design measure $\xi = n^{-1}\Sigma_{i=1}^{n} I(\mathbf{x}=\mathbf{x}_i)$ as well as on w, f and g. Formally,

$$\mathrm{IMSPE}(f, g, w, \xi) = n\int_S E\{[\hat{Y}(\mathbf{x}) - E(Y|\mathbf{x})]^2\}\,\mathrm{d}\mathbf{x},$$

$$\mathrm{IMSEE}(f, g, w, \xi) = n\int_T E\{[\hat{Y}(\mathbf{x}) - E(Y|\mathbf{x})]^2\}\,\mathrm{d}\mathbf{x}.$$

There is a sizeable literature concerning regression designs for a possibly misspecified linear response. Such designs for homoscedastic errors have been studied by Box and Draper (1959), Huber (1975) and Wiens (1992). Designs for prediction with as well possible heteroscedasticity were obtained by Wiens (1998). For extrapolation with homoscedastic errors see Draper and Herzberg (1973), Huber (1975), Lawless (1984) and Spruill (1984). In these studies, the goal was extrapolation to one fixed point on or outside the boundary of the design space. Robust designs for extrapolation with possible heteroscedasticity were obtained by Fang and Wiens (1999). Designs for extrapolation to one point outside the design space were studied by Dette and Wong (1996), whose extrapolation designs for polynomial responses are robust against misspecification of the degree of the polynomial, and more recently by Wiens and Xu (2005).

For nonlinear regression, Atkinson and Haines (1996) and Ford et al. (1989) present various static and sequential designs for nonlinear models without the consideration of model uncertainty. Sinha and Wiens (2002) also employ notions of robustness in the construction of sequential designs for the nonlinear model. In addition, Wiens and Xu (2005) discuss the construction of robust designs for a possibly misspecified nonlinear model and for extrapolation of a regression response to one point outside of the design space. The current work goes beyond that of Wiens and Xu (2005) in that we deal with both prediction and extrapolation and, in the latter case, we allow the extrapolation space T to have nonzero measure. We go beyond Fang and Wiens (1999) in treating nonlinear models. The major advance, though, is perhaps our treatment of essentially unrestricted regressors $\mathbf{z}(\mathbf{x})$. Explicit designs in almost all problems involving misspecified regressors were hitherto restricted to cases in which $\mathbf{z}(\mathbf{x})$ was well structured—e.g. straight line regression ($\mathbf{z}(x) = (1, x)^{\mathrm{T}}$), polynomial regression, or multiple regression without interactions on a spherical design space ($\mathbf{z}(\mathbf{x}) = (1, \mathbf{x}^{\mathrm{T}})^{\mathrm{T}}$, $\|\mathbf{x}\| \leqslant$ const.). The improvements in the current work are made possible by our adaptation of recent results of Shi et al. (2003), henceforth referred to as SYZ.

SYZ investigated the analytical form of minimax designs for prediction problems when the function f was an unknown member of the class

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

In our terminology, they considered the case of approximate linearity—$h(\mu)=\mu$—and homoscedasticity $g = \mathbf{1}$, where $\mathbf{1}(\mathbf{x}) = \mathbf{x}$. The orthogonality condition in $\mathscr{F}$ ensures that the parameter $\boldsymbol{\theta}$ is uniquely defined in model (1). The second condition assures that overall f is not too large.

The class $\mathscr{F}$ is sufficiently rich that any 'design' with finite maximum loss must have a density, and thus must be approximated to make it implementable. Approximation methods are discussed in Heo et al. (2001). These can, for instance, take the form of choosing the design points so as to obtain agreement between the $(i-1)/(n-1)$-quantiles ($i = 1, \ldots, n$) of the empirical and theoretical design measures, or between the moments to a sufficiently high order.

SYZ show that the minimax design densities are of the form

$$m(\mathbf{x}) = \left[\frac{\mathbf{z}^{\mathrm{T}}(\mathbf{x})\mathbf{P}\mathbf{z}(\mathbf{x}) + d}{\mathbf{z}^{\mathrm{T}}(\mathbf{x})\mathbf{Q}\mathbf{z}(\mathbf{x})}\right]^{+}$$

for almost all $\mathbf{x} \in S$, where $c^+ = \max(c, 0)$, for suitable constant symmetric matrices $\mathbf{P}$, $\mathbf{Q}$ and a constant d. These constants may then be determined numerically.

In this article we extend SYZ so as to obtain robust designs for extrapolation and prediction, assuming that the regression response is as at (2) and that the errors may be heteroscedastic. If the function h in (2) is not the identity then our designs are only locally optimal. They are, however, still of substantial practical interest—see reasons for this as listed in Ford et al. (1992) and restated in Ford et al. (1989). One typical reason is that where sequential designs can be carried out in batches, the design for the next batch might be a locally optimal design based on the estimates obtained from the previous batch. Allowing for uncertainty in our best guess at a local parameter, we adopt the approach introduced in Wiens and Xu (2005) to find 'locally most robust' designs which are minimax with respect to a region containing the initial parameters.

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

(P1) Ordinary least squares (OLS) estimation with homoscedasticity: determine designs to minimize the maximum value, over f, of IMSEE$(f, \mathbf{1}, \mathbf{1}, \xi)$.
(P2) OLS with heteroscedasticity: determine designs to minimize the maximum value, over f and g, of IMSPE$(f, g, \mathbf{1}, \xi)$.
(P3) OLS with heteroscedasticity: determine designs to minimize the maximum value, over f and g, of IMSEE$(f, g, \mathbf{1}, \xi)$.
(P4) Weighted least squares (WLS) estimation with heteroscedasticity: determine designs and weights to minimize the maximum value, over f and g, of IMSPE(f, g, w, ξ).
(P5) WLS with heteroscedasticity: determine designs and weights to minimize the maximum value, over f and g, of IMSEE(f, g, w, ξ).

The rest of this article is organized as follows. The designs for P1 are provided in Section 3. Those for P2 and P3 are given in Section 4. The designs and weights which constitute the solutions to problems P4 and P5 are given in Section 5. Some mathematical preliminaries are detailed in Section 2. We present several examples in Section 6, and conclude with a few remarks in Section 7. Derivations are provided in the Appendix.

2. Preliminaries and notation

The regression models discussed in this paper are very similar to those in Wiens and Xu (2005), except that we consider the prediction case as well and allow the extrapolation space to be any space, of positive Lebesgue measure, outside the design space. For the reader's convenience, we briefly describe this model here.

We assume that the contaminant $f(\cdot)$ is an unknown member of

$$\mathscr{F} = \left\{ f \,\middle|\, \int_S f^2(\mathbf{x})\,\mathrm{d}\mathbf{x} \leqslant \eta_S^2 < \infty, \int_T f^2(\mathbf{x})\,\mathrm{d}\mathbf{x} \leqslant \eta_T^2 < \infty, \int_S \tilde{\mathbf{z}}(\mathbf{x}) f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \mathbf{0} \right\}, \tag{3}$$

where $\mu = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}(\mathbf{x})$, $\tilde{\mathbf{z}}(\mathbf{x}) = (\mathrm{d}h/\mathrm{d}\mu|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0})\mathbf{z}(\mathbf{x})$ and η_S, η_T are positive constants. For prediction problems ($T = S$) the second condition in (3) merges into the first. The last condition is required in order that $\boldsymbol{\theta}_0$ can be uniquely defined, and in fact arises through the definition

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

together with

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

Where possible, we drop the subscript on f.

The observations Y_i, although uncorrelated with mean $h(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i)) + n^{-1/2} f(\mathbf{x}_i)$, are possibly heteroscedastic with

$$var\{Y(\mathbf{x}_i)\} = \sigma^2 g(\mathbf{x}_i) \tag{4}$$

for a function g satisfying conditions given in Section 4.

For extrapolation problems, the only assumptions made about T are that it is disjoint from S and has nonzero Lebesgue measure. To ensure the nonsingularity of a number of relevant matrices, we assume that the design and extrapolation spaces satisfy

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

We make use of the following matrices and vectors:

$$\begin{aligned} \mathbf{A}_S &= \textstyle\int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\,\mathrm{d}\mathbf{x}, & \mathbf{A}_T &= \textstyle\int_T \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\,\mathrm{d}\mathbf{x}, \\ \mathbf{B} &= \textstyle\int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})w(\mathbf{x})\xi(\mathrm{d}\mathbf{x}), & \mathbf{D} &= \textstyle\int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})w^2(\mathbf{x})g(\mathbf{x})\xi(\mathrm{d}\mathbf{x}), \\ \mathbf{b}_{f,S} &= \textstyle\int_S \tilde{\mathbf{z}}(\mathbf{x})f(\mathbf{x})w(\mathbf{x})\xi(\mathrm{d}\mathbf{x}), & \mathbf{b}_{f,T} &= \textstyle\int_T \tilde{\mathbf{z}}(\mathbf{x})f(\mathbf{x})\,\mathrm{d}\mathbf{x}. \end{aligned}$$

It follows from (A) that $\mathbf{A}_S$ is nonsingular and that $\mathbf{B}$ is nonsingular as well if, as is assumed below, ξ is absolutely continuous. The LSE of $\boldsymbol{\theta}_0$ is

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

The information matrix is

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

and the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is

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

For prediction, the loss function IMSPE splits into terms due to bias, variance and model misspecification:

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

where the integrated bias (IPB) and integrated variance (IPV) are

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

and

$$\mathrm{IPV}(g, w, \xi) = n\int_S VAR(\hat{Y}(\mathbf{x}))\,\mathrm{d}\mathbf{x} = n\int_S VAR(h(\hat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{z}(\mathbf{x})))\,\mathrm{d}\mathbf{x}.$$

Asymptotically,

$$\begin{aligned} \mathrm{IPB}(f, w, \xi) &= \mathbf{b}_{f,S}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{A}_S\mathbf{B}^{-1}\mathbf{b}_{f,S}, \\ \mathrm{IPV}(g, w, \xi) &= \sigma^2\,\mathrm{tr}(\mathbf{B}^{-1}\mathbf{A}_S\mathbf{B}^{-1}\mathbf{D}). \end{aligned}$$

(The second term in IPB vanishes asymptotically by virtue of the orthogonality condition in the definition of $\mathscr{F}$.)

For extrapolation, the loss function IMSEE decomposes in a similar fashion:

$$\mathrm{IMSEE}(f, g, w, \xi) = n\int_T E\{[\hat{Y}(\mathbf{x}) - E(Y|\mathbf{x})]^2\}\,\mathrm{d}\mathbf{x} = \mathrm{IEB}(f, w, \xi) + \mathrm{IEV}(g, w, \xi) + \int_T f^2(\mathbf{x})\,\mathrm{d}\mathbf{x},$$

where, asymptotically,

$$\mathrm{IEB}(f, w, \xi) = \mathbf{b}_{f,S}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\mathbf{b}_{f,S} - 2\mathbf{b}_{f,T}\mathbf{B}^{-1}\mathbf{b}_{f,S},$$

$$\mathrm{IEV}(g, w, \xi) = \sigma^2\,\mathrm{tr}(\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\mathbf{D}).$$

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

$$\int_S \frac{m(\mathbf{x})}{w(\mathbf{x})}\,\mathrm{d}\mathbf{x} = 1 \tag{5}$$

and

$$\mathbf{B} = \int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})m(\mathbf{x})\,\mathrm{d}\mathbf{x},$$

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

From the definitions of $\mathbf{B}$, $\mathbf{b}_{f,S}$ and $\mathbf{b}_{f,T}$, we notice that $\mathrm{IPB}(f, w, \xi)$ and $\mathrm{IEB}(f, w, \xi)$ rely on (w, ξ) only through m and $\mathrm{IPV}(g, w, \xi)$ and $\mathrm{IEV}(g, w, \xi)$ through m and w. Hence, we can optimize over m and w subject to (5) rather than over k and w.

Although the IEB may be negative,

$$\mathrm{IEB} + \int_T f^2(\mathbf{x})\,\mathrm{d}\mathbf{x} = n\int_T \{E[h(\hat{\boldsymbol{\theta}}^{\mathrm{T}}\tilde{\mathbf{z}}(\mathbf{x})) - h(\boldsymbol{\theta}_0^{\mathrm{T}}\tilde{\mathbf{z}}(\mathbf{x})) - n^{-1/2}f(\mathbf{x})]\}^2\,\mathrm{d}\mathbf{x} \geqslant 0.$$

We define $r_{T,S} = \eta_T/\eta_S$, reflecting the relative amounts of model response uncertainty in the extrapolation and design spaces, and $v = \sigma^2/\eta_S^2$, representing the relative importance of variance versus bias. We remark that for prediction our results depend on the unknown parameters only through v and $\boldsymbol{\theta}_0$, while for extrapolation they depend on the parameters only through $r_{T,S}$, v and $\boldsymbol{\theta}_0$. In the special case $h(\mu) = \mu$, the results are independent of $\boldsymbol{\theta}_0$.

We also require the definitions $\mathbf{K} = \int_S \tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})m^2(\mathbf{x})\,\mathrm{d}\mathbf{x}$, $\mathbf{G} = \mathbf{K} - \mathbf{B}\mathbf{A}_S^{-1}\mathbf{B}$, $\mathbf{H}_S = \mathbf{B}^{-1}\mathbf{A}_S\mathbf{B}^{-1}$, and $\mathbf{H}_T = \mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}$.

In the next three sections, we will exhibit solutions to P1–P5.

3. Optimal extrapolation designs with homoscedasticity: solutions to P1

SYZ provide the form of the minimax density for prediction when $h = \mathbf{1}$. In this section, we extend this result to extrapolation and to a general h.

Denote the largest eigenvalue of a matrix $\mathbf{X}$ by $\lambda_{\max}(\mathbf{X})$. As at Theorem 2.1(a) in Fang and Wiens (1999), the maximum extrapolation bias is

$$\sup_{f\in\mathscr{F}} \mathrm{IEB}(f, \mathbf{1}, \xi) = \eta_S^2[(\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)} + r_{T,S})^2 - r_{T,S}^2] \geqslant 0.$$

Therefore, the maximum IMSEE is

$$\sup_{f\in\mathscr{F}} \mathrm{IMSEE}(f, \mathbf{1}, \mathbf{1}, m) = \eta_S^2\left[(\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)} + r_{T,S})^2 + v\int_S \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_T\tilde{\mathbf{z}}(\mathbf{x})m(\mathbf{x})\,\mathrm{d}\mathbf{x}\right]$$

$$= \eta_S^2\left[(\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)} + r_{T,S})^2 + v\,\mathrm{tr}(\mathbf{B}^{-1}\mathbf{A}_T)\right]. \tag{6}$$

A minimax design is one for which the density m minimizes (6). This is an optimization problem with an objective function involving a generally nonsmooth function $\lambda_{\max}$. Employing nonsmooth optimization theory (Clarke, 1983; see SYZ for a useful review), we obtain the following result.

Theorem 1. *The minimax design density for extrapolation, when the variances are homogeneous, is of the form*

$$m(\mathbf{x}) = \left[\frac{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{P}\tilde{\mathbf{z}}(\mathbf{x}) + d}{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{Q}\tilde{\mathbf{z}}(\mathbf{x})}\right]^{+} \tag{7}$$

for almost all $\mathbf{x} \in S$, *for constant symmetric matrices* $\mathbf{P}, \mathbf{Q}(\geqslant \mathbf{0})$ *and a constant d. The constants minimize* (6) *and satisfy* $\int_S m(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$.

Remarks. 1. As in SYZ, in the examples for linear regression in this article, we only consider symmetric densities when the structure of the design and extrapolation spaces make this appropriate.

2. The symmetric—in each component of $\mathbf{x}$—minimax density has the form exhibited in Theorem 1 but with the odd functions of these components vanishing. The proof of this is very similar to the proof in Shi (2002) for linear regression.

Example 3.1. For the regression model

$$Y = \theta_0 + \theta_1 x + \theta_2 x^2 + f(x) + \varepsilon, \quad x \in [-a, a]$$

with symmetric extrapolation space $[-r_2, -r_1) \cup (r_1, r_2]$ for $0 < a \leqslant r_1 < r_2$, it is reasonable to restrict to symmetric designs. According to Theorem 1, the symmetric optimal design for this model with homoscedasticity is of the form

$$m(x) = \left(\frac{a_1 + a_2x^2 + a_3x^4}{a_4 + a_5x^2 + a_6x^4}\right)^{+}, \tag{8}$$

where a_4 and a_6 are nonnegative. Some computations for this case are shown in Example 6.1.

Example 3.2. For the linear regression model with two interacting regressors

$$Y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_1 x_2 + f(x_1, x_2) + \varepsilon,$$

with $S = [-a, a] \times [-a, a]$ and $T = [-r, r] \times [-r, r] \backslash S$ $(r > a)$, the minimax designs for prediction were studied by Adewale (2002) who states that the symmetric, exchangeable minimax density is given by

$$m(x_1, x_2) = \left(\frac{a + b(x_1^2 + x_2^2) + cx_1^2x_2^2}{a' + b'(x_1^2 + x_2^2) + c'x_1^2x_2^2}\right)^{+}.$$

From Theorem 1, the minimax symmetric and exchangeable density for extrapolation is also of this form.

Example 3.3. For the nonlinear regression model

$$Y = \mathrm{e}^{\theta_0 + \theta_1 x} + f(x) + \varepsilon, \tag{9}$$

for which $h(\mu) = \mathrm{e}^{\mu}$, we take $S = [0, 1]$ and $T = (1, r]$. The locally most robust extrapolation design density is given by

$$m(x) = \left(\frac{\mathrm{e}^{2\theta_1 x}(a_1 + b_1 x + c_1 x^2) + d}{\mathrm{e}^{2\theta_1 x}(a_2 + b_2 x + c_2 x^2)}\right)^{+},$$

where $a_2 \geqslant 0$, $c_2 \geqslant 0$ and $a_1, b_1, c_1, a_2, b_2, c_2$ and d chosen in order to minimize (6) subject to $\int_0^1 m(x)\,\mathrm{d}x = 1$. The dependence of the design on θ_1 is an issue which will be addressed in Example 6.2.

4. Optimal prediction and extrapolation designs with heteroscedasticity for OLS: solutions to P2 and P3

In this and the next section we construct designs which are robust against heteroscedasticity as well as against departures from the fitted response. The heteroscedasticity is governed by $g(\cdot)$—recall (4)—which is assumed to belong to

$$\mathscr{G} = \left\{ g \,\middle|\, \int_S g^2(\mathbf{x})\,\mathrm{d}\mathbf{x} \leqslant \Omega^{-1} = \int_S \mathrm{d}\mathbf{x} < \infty \right\}. \tag{10}$$

In (10), the equality condition is equivalent to defining

$$\sigma^2 = \sup_g \left[\int_S var^2\{\varepsilon(\mathbf{x})\}\Omega\,\mathrm{d}\mathbf{x} \right]^{1/2}.$$

As at Theorem 1(c) in Wiens (1998) and Theorem 2.1(c) of Fang and Wiens (1999), for OLS the maximum integrated mean square prediction error and extrapolation error are

$$\sup_{f\in\mathscr{F}, g\in\mathscr{G}} \mathrm{IMSPE}(f, g, \mathbf{1}, m) = \eta_S^2 \left\{ \begin{array}{l} \lambda_{\max}(\mathbf{K}\mathbf{H}_S) \\ \quad +v\Omega^{-1/2}\left[\int_S \{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_S\tilde{\mathbf{z}}(\mathbf{x})m(\mathbf{x})\}^2\,\mathrm{d}\mathbf{x}\right]^{1/2} \end{array} \right\}, \tag{11}$$

$$\sup_{f\in\mathscr{F}, g\in\mathscr{G}} \mathrm{IMSEE}(f, g, \mathbf{1}, m) = \eta_S^2 \left\{ \begin{array}{l} (\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)} + r_{T,S})^2 \\ \quad +v\Omega^{-1/2}\left[\int_S \{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_T\tilde{\mathbf{z}}(\mathbf{x})m(\mathbf{x})\}^2\,\mathrm{d}\mathbf{x}\right]^{1/2} \end{array} \right\}, \tag{12}$$

respectively. Therefore problem P2 requires finding a density $m(\cdot)$ which minimizes (11), whereas P3 requires finding a density which minimizes (12).

Theorem 2. *The minimax design densities for both prediction and extrapolation with OLS estimation, when the variances are possibly heterogeneous, have the form*

$$m(\mathbf{x}) = \left[\frac{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{P}\tilde{\mathbf{z}}(\mathbf{x}) + d}{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{Q}\tilde{\mathbf{z}}(\mathbf{x}) + \{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{U}\tilde{\mathbf{z}}(\mathbf{x})\}^2} \right]^+ \tag{13}$$

for almost all $\mathbf{x} \in S$, *for constant symmetric matrices* $\mathbf{P}$, $\mathbf{Q}(\geqslant \mathbf{0})$, $\mathbf{U}(> \mathbf{0})$ *and a constant d such that* (1) $\int_S m(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$ *and* (2) *for prediction*, (11) *is minimized*, *while for extrapolation* (12) *is minimized.*

Example 4.1. For the simple linear regression model

$$Y = \theta_0 + \theta_1 x + f(x) + \varepsilon, \tag{14}$$

with $S = [-1, 1]$, the minimax prediction design was studied by Wiens (1998). It was shown there that the minimax symmetric density is given by

$$m(x) = \left(\frac{a + bx^2}{1 + cx^2 + dx^4} \right)^+, \tag{15}$$

a form which now follows as well from Theorem 2. Similarly, for extrapolation, Fang and Wiens (1999) derive the form (15). More generally, for OLS in the multiple linear regression model

$$Y = \theta_0 + \sum_{j=1}^{p-1} \theta_j x_j + f(\mathbf{x}) + \varepsilon,$$

with S being a unit hypersphere centred at the origin and $T = \{\mathbf{x} | 1 < \|\mathbf{x}\| \leqslant r\}$, Fang and Wiens (1999) obtained conditions under which the minimax symmetric extrapolation design density would be given by

$$m(\mathbf{x}) = \left(\frac{a + b\|\mathbf{x}\|^2}{c + d\|\mathbf{x}\|^2 + e\|\mathbf{x}\|^4} \right)^+.$$

This form now follows, without conditions, from Theorem 2 and Remark 2 in Section 3.

Example 4.2. For the nonlinear model (9) it follows from Theorem 2 that the locally optimal design density for both prediction and extrapolation is of the form

$$m(x)=\left(\frac{e^{2\theta_1 x}(a_1+b_1x+c_1x^2)+d}{e^{2\theta_1 x}[(a_2+b_2x+c_2x^2)+e^{2\theta_1 x}(a_3+b_3x+c_3x^2)^2]}\right)^+, \tag{16}$$

where $a_2\geqslant 0$, $c_2\geqslant 0$, $a_3>0$ and $c_3>0$. When $\theta_1=0$, (16) can be reduced to

$$m^*(\mathbf{x})=\left[\frac{a_1+a_2x+a_3x^2}{1+a_5x+a_6x^2+a_7x^3+a_8x^4}\right]^+,$$

where a_6 and a_8 are positive. The computation of our designs for this model are detailed in Example 6.2.

5. Optimal prediction and extrapolation designs with heteroscedasticity for WLS: solutions to P4 and P5

In this section we propose to estimate $\boldsymbol{\theta}$ by WLS, and again consider both prediction and extrapolation problems. For prediction we proceed as in Wiens (1998) and obtain

$$\sup_{f\in\mathscr{F},g\in\mathscr{G}}\mathrm{IMSPE}(f,g,w,m)=\eta_S^2\left\{\lambda_{\max}(\mathbf{KH}_S)+v\Omega^{-1/2}\left[\int_S\{w(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_S\tilde{\mathbf{z}}(\mathbf{x})m(\mathbf{x})\}^2\,\mathrm{d}\mathbf{x}\right]^{1/2}\right\}. \tag{17}$$

The weights minimizing (17) for fixed $m(\mathbf{x})$, subject to $\int_S(m(\mathbf{x})/w(\mathbf{x}))\,\mathrm{d}\mathbf{x}=1$, are, in terms of

$$\alpha_{S,m}=\int_S[\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_S\tilde{\mathbf{z}}(\mathbf{x})m^2(\mathbf{x})]^{2/3}\,\mathrm{d}\mathbf{x},$$

given by

$$w_{S,m}(\mathbf{x})=\alpha_{S,m}[\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_S\tilde{\mathbf{z}}(\mathbf{x})\}^2m(\mathbf{x})]^{-1/3}I\{m(\mathbf{x})>0\}. \tag{18}$$

Then

$$\min_w\sup_{f\in\mathscr{F},g\in\mathscr{G}}\mathrm{IMSPE}(f,g,w,m)=\eta_S^2\{\lambda_{\max}(\mathbf{KH}_S)+v\Omega^{-1/2}\alpha_{S,m}^{3/2}\} \tag{19}$$

and problem P4 becomes that of finding a density $m^*(\mathbf{x})$ which minimizes (19). Then the weights $w_{S,m^*}(\mathbf{x})$ obtained from (18) and the design density

$$k_*(\mathbf{x})=\frac{m^*(\mathbf{x})}{w_{S,m^*}(\mathbf{x})}=\alpha_{S,m^*}^{-1}\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_S\tilde{\mathbf{z}}(\mathbf{x})m^{*2}(\mathbf{x})\}^{2/3}$$

are optimal for WLS prediction.

For extrapolation we follow Fang and Wiens (1999) and obtain

$$\sup_{f\in\mathscr{F},g\in\mathscr{G}}\mathrm{IMSEE}(f,g,w,m)=\eta_S^2\left\{(\sqrt{\lambda_{\max}(\mathbf{GH}_T)}+r_{T,S})^2+v\Omega^{-1/2}\left[\int_S\{w(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_T\tilde{\mathbf{z}}(\mathbf{x})m(\mathbf{x})\}^2\,\mathrm{d}\mathbf{x}\right]^{1/2}\right\}.$$

The minimizing weights are given by

$$w_{T,m}(\mathbf{x})=\alpha_{T,m}[\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_T\tilde{\mathbf{z}}(\mathbf{x})\}^2m(\mathbf{x})]^{-1/3}I\{m(\mathbf{x})>0\}, \tag{20}$$

with

$$\min_w\sup_{f\in\mathscr{F},g\in\mathscr{G}}\mathrm{IMSEE}(f,g,m)=\eta_S^2\{(\sqrt{\lambda_{\max}(\mathbf{GH}_T)}+r_{T,S})^2+v\Omega^{-1/2}\alpha_{T,m}^{3/2}\} \tag{21}$$

and to solve P5 we seek a density $m^*(\mathbf{x})$ which minimizes (21). Then the weights $w_{T,m^*}(\mathbf{x})$ obtained from (20) and the design density

$$k_*(\mathbf{x}) = \frac{m^*(\mathbf{x})}{w_{T,m^*}(\mathbf{x})} = \alpha_{T,m^*}^{-1}\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_T\tilde{\mathbf{z}}(\mathbf{x})m^{*2}(\mathbf{x})\}^{2/3} \tag{22}$$

are optimal for WLS extrapolation.

The following theorem provides the form of $m^*(\mathbf{x})$ for both prediction and extrapolation.

Theorem 3. *The minimax densities $m^*(\mathbf{x})$ for both prediction and extrapolation with WLS estimation, when the variances are possibly heterogeneous, are of the form*

$$m^*(\mathbf{x}) = \left[\frac{c(\mathbf{x}) - k(\mathbf{x})}{b(\mathbf{x})}\right]^+, \tag{23}$$

where, for constant symmetric matrices $\mathbf{P}, \mathbf{Q}(\geqslant \mathbf{0}), \mathbf{U}(> \mathbf{0})$ and a constant d we have $b(\mathbf{x}) = \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{Q}\tilde{\mathbf{z}}(\mathbf{x})$, $c(\mathbf{x}) = \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{P}\tilde{\mathbf{z}}(\mathbf{x}) + d$ and

$$k^3 + \frac{a^3}{b}k - \frac{a^3 c}{b} = 0,$$

with $a(\mathbf{x}) = \{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{U}\tilde{\mathbf{z}}(\mathbf{x})\}^{2/3}$. Explicitly,

$$k = a\left[\left\{\frac{c}{2b} + \sqrt{\left(\frac{c}{2b}\right)^2 + \left(\frac{a}{3b}\right)^3}\right\}^{1/3} + \left\{\frac{c}{2b} - \sqrt{\left(\frac{c}{2b}\right)^2 + \left(\frac{a}{3b}\right)^3}\right\}^{1/3}\right]. \tag{24}$$

The constants satisfy (1) $\int_S m(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$ *and* (2) *minimize* (19) *for prediction*, (21) *for extrapolation.*

Example 5.1. For the simple linear regression model (14) with $S = [-1, 1]$ and $T = \{x | 1 < |x| \leqslant r\}$ we obtain (23) with

$$\begin{aligned} c(x) &= a_1 + a_2 x^2, \\ b(x) &= a_3 + a_4 x^2, \\ a(x) &= (a_5 + a_6 x^2)^{2/3}, \end{aligned}$$

where $a_3 \geqslant 0$, $a_4 \geqslant 0$, $a_3^2 + a_4^2 > 0$, $a_5 > 0$ and $a_6 > 0$ are determined as in the statement of Theorem 3. The minimax weights are obtained from (18) and (20) with

$$\begin{aligned} \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_S\tilde{\mathbf{z}}(\mathbf{x}) &= 2 + \tfrac{2}{3}x^2\left(\int_{-1}^{1} x^2 m(x)\,\mathrm{d}x\right)^{-2}, \\ \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_T\tilde{\mathbf{z}}(\mathbf{x}) &= 2(r-1) + \tfrac{2}{3}(r^3-1)x^2\left(\int_{-1}^{1} x^2 m(x)\,\mathrm{d}x\right)^{-2}. \end{aligned}$$

Example 5.2. For the nonlinear model (9) with $S = [0, 1]$ and $T = (1, r]$ $(r > 1)$ we attain (23) with

$$\begin{aligned} c(x) &= a_0 + \mathrm{e}^{2\theta_1 x}(a_1 + a_2 x + a_3 x^2), \\ b(x) &= \mathrm{e}^{2\theta_1 x}(a_4 + a_5 x + a_6 x^2), \\ a(x) &= [\mathrm{e}^{2\theta_1 x}(a_7 + a_8 x + a_9 x^2)]^{2/3}, \end{aligned}$$

where $a_4 \geqslant 0$, $a_6 \geqslant 0$, $a_4^2 + a_6^2 > 0$, $a_7 > 0$ and $a_9 > 0$ are determined as in Theorem 3. Note that the term $\mathrm{e}^{2\theta_0}$ has been absorbed into $a_1, \ldots, a_9$. The minimax weights are derived from (18) and (20) with

$$\begin{aligned} \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_S\tilde{\mathbf{z}}(\mathbf{x}) &= (u_1 u_3 - u_2^2)^{-2}\mathrm{e}^{2\theta_1 x}\phi(x; s_1, s_2, s_3, u_1, u_2, u_3), \\ \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_T\tilde{\mathbf{z}}(\mathbf{x}) &= (u_1 u_3 - u_2^2)^{-2}\mathrm{e}^{2\theta_1 x}\phi(x; t_1, t_2, t_3, u_1, u_2, u_3), \end{aligned}$$

where

$$\phi(x; s_1, s_2, s_3, u_1, u_2, u_3) = (u_3^2 s_1 - 2u_2u_3s_2 + u_2^2 s_3) + 2(u_1u_3s_2 - u_1u_2s_3 - u_2u_3s_1 + u_2^2 s_2)x + (u_1^2 s_3 - 2u_1u_2s_2 + u_2^2 s_1)x^2,$$

$$s_1 = \frac{e^{2\theta_1} - 1}{2\theta_1}, \quad s_2 = \frac{e^{2\theta_1} - s_1}{2\theta_1}, \quad s_3 = \frac{e^{2\theta_1} - 2s_2}{2\theta_1},$$

$$t_1 = \frac{e^{2\theta_1 r} - e^{2\theta_1}}{2\theta_1}, \quad t_2 = \frac{re^{2\theta_1 r} - e^{2\theta_1} - t_1}{2\theta_1}, \quad t_3 = \frac{r^2 e^{2\theta_1 r} - e^{2\theta_1} - 2t_2}{2\theta_1},$$

$$u_1 = \int_0^1 e^{2\theta_1 x} m(x)\,dx, \quad u_2 = \int_0^1 x e^{2\theta_1 x} m(x)\,dx, \quad u_3 = \int_0^1 x^2 e^{2\theta_1 x} m(x)\,dx.$$

6. Computations and examples

Example 6.1. Recall Example 3.1 and (8). We take $r_1 = 1$ and denote r_2 by r. If either of a_4 or a_6 is nonzero, we may take it to be unity. We take $a_4 = 1$ and $r_{T,S} = 1$. Some numerical values of the constants are shown in Table 1. Fig. 1 gives

Table 1
Coefficient values for the density (8) with $a_4 = 1$ for Example 6.1

r	v	a_1	a_2	a_3	a_5	a_6	Loss
1.5	0.1	0.377	78.22	273.86	278.32	149.87	6.79
	0.25	1.16	−6.45	289.98	217.64	27.96	9.50
	1	1.16	−37.05	70.72	−2.60	18.62	19.70
	5	0.804	−13.31	16.85	−3.98	4.18	61.06
	10	0.923	−20.53	24.57	−3.99	3.98	108.49
	100	1.65	−124.34	132.51	−3.50	3.06	853.72
5	0.1	0.524	−1.04	1.55	−1.10	1.59	1302.54
	0.25	0.654	−4.17	5.71	−5.02	6.30	2544.55
	1	0.822	−6.52	8.38	−4.62	5.33	8508.54
	5	1.21	−16.62	19.64	−4.28	4.59	34294.25
	10	1.47	−28.05	32.05	−4.19	4.39	63387.51
	100	2.98	−214.47	226.23	−4.02	4.04	534222.2

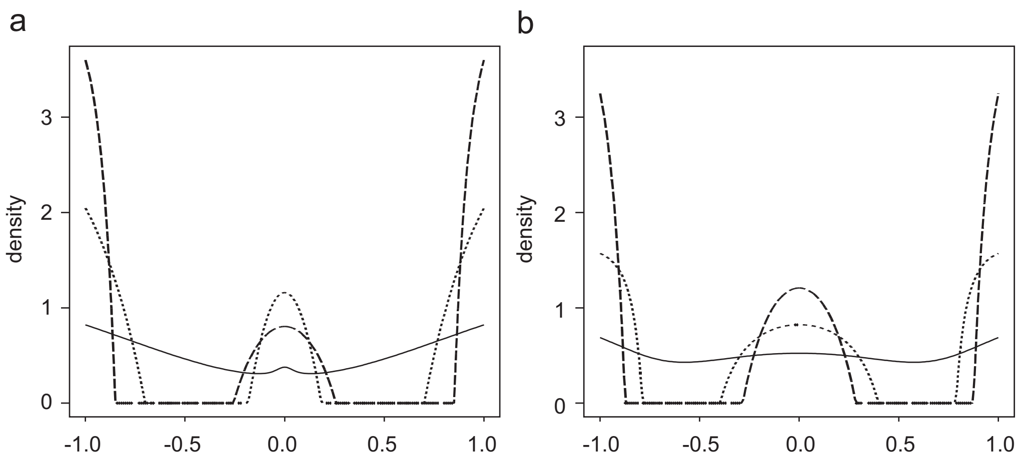

Fig. 1. Minimax extrapolation densities $m(x) = ((a_1 + a_2x^2 + a_3x^4)/(1 + a_5x^2 + a_6x^4))^+$ in Example 6.1. (a) $r = 1.5$ and (b) $r = 5$. Each plot uses three values of v: $v = 0.1$ (solid line), $v = 1$ (dotted line), $v = 5$ (broken line).

Table 2
Coefficient values for the density (16) in Example 6.2 with $a_2 = 1$ and $\theta_1 = 1$

v	a_1	b_1	c_1	d	b_2	c_2	a_3	b_3	c_3
0.5	−513.21	495.36	6.99	782.40	3.44	46.96	16.07	−39.74	25.31
1	932.34	−2428.57	1892.85	−322.54	12.23	106.62	25.44	−60.19	38.81
5	2044.47	−4883.06	3303.49	−816.61	7.95	25.45	35.90	−81.08	50.00

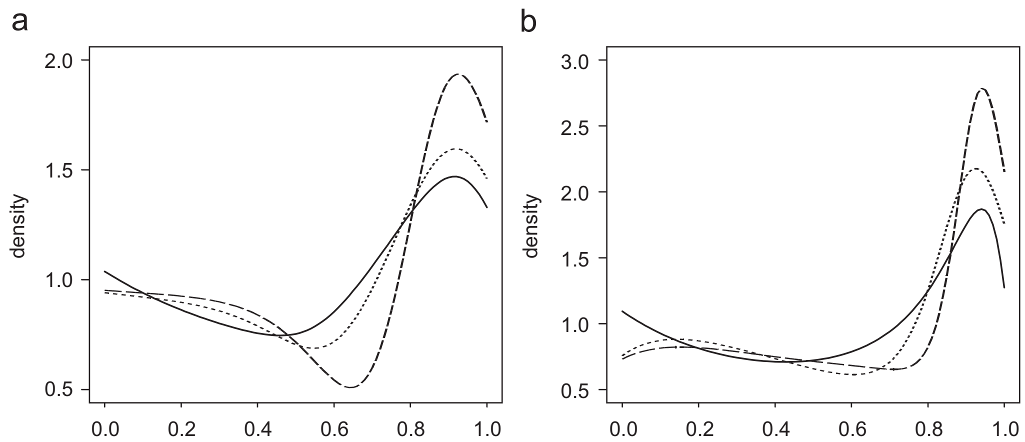

Fig. 2. Optimal minimax design densities $m(x) = ((e^{2\theta_1^* x}(a_1 + b_1 x + c_1 x^2) + d)/(e^{2\theta_1^* x}[(1 + b_2 x + c_2 x^2) + e^{2\theta_1^* x}(a_3 + b_3 x + c_3 x^2)^2]))^+$ in Example 6.2. (a) Locally optimal design densities for $\theta_1^* = 1$ and (b) locally most robust design densities for $\theta_1^* = \theta_1^{LF}$ in [0, 2]. Each plot uses three values of v: $v = 0.5$ (solid line), $v = 1$ (dotted line), $v = 5$ (broken line).

Table 3
Coefficient values for the locally most robust density (16)

v	a_1	b_1	c_1	d	b_2	c_2	a_3	b_3	c_3
0.5	24022.10	−28233.18	4874.55	3336.73	1.78	20.02	158.10	−291.20	136.26
1	92029.01	−217292.1	138815.1	−69222.57	1.34	8.32	173.63	−362.91	200.60
5	122521.4	−276528.6	161973.8	−85323.74	9.06	105.07	225.56	−483.99	265.67

plots of the minimax extrapolation densities for varying r and v. The designs can be roughly described as replacing those points masses at −1, 1 and 0 in the variance minimizing designs by more or less uniformly distributed clusters in neighbourhoods of these points. Decreasing v results in more uniform designs. A larger r (wider extrapolation region) results in more uniformity as well, especially in the central region.

Example 6.2. Recall Example 4.2 and the nonlinear model (9) with possible heteroscedasticity. The locally optimal design density for prediction is given by (16). See Table 2 for the numerical values of the constants in (16) and Fig. 2(a) for plots. Here we have taken $a_2 = 1$ and $\theta_1 = 1$.

These designs are only locally optimal since they depend on the value of θ_1. To deal with this, we obtain 'locally most robust' designs as in Wiens and Xu (2005). For this, we take a further maximum of the loss as θ_1 varies over some interval I, and determine the coefficients of $m(x)$ so as to minimize this maximum loss. For $I = [0, 2]$, the locally most robust designs are detailed in Table 3 for varying v. In each case, we found that the least favourable θ_1 within I, say θ_1^{LF}, is 2. See Fig. 2(b) for plots. Although, as pointed out in Silvey (1980), local designs tailored for optimality at a least favourable parameter value are sometimes inefficient at distant points, it has been our experience that the designs constructed here do not exhibit a strong dependence on θ_1.

Table 4
Coefficient values for the locally optimal product (23) of density and weights in Example 6.3

v	a_0	a_1	a_2	a_3	a_5	a_6	a_7	a_8	a_9
0.5	163.19	582.58	−369.48	211.80	6.18	26.47	18978.42	165.78	1122.10
1	134.09	2710.50	−2352.19	1737.83	11.92	84.98	138644.97	4451.39	65374.78
2	−3269.85	10678.54	−12184.94	9087.93	13.45	81.59	581177.64	20860.88	701666.09

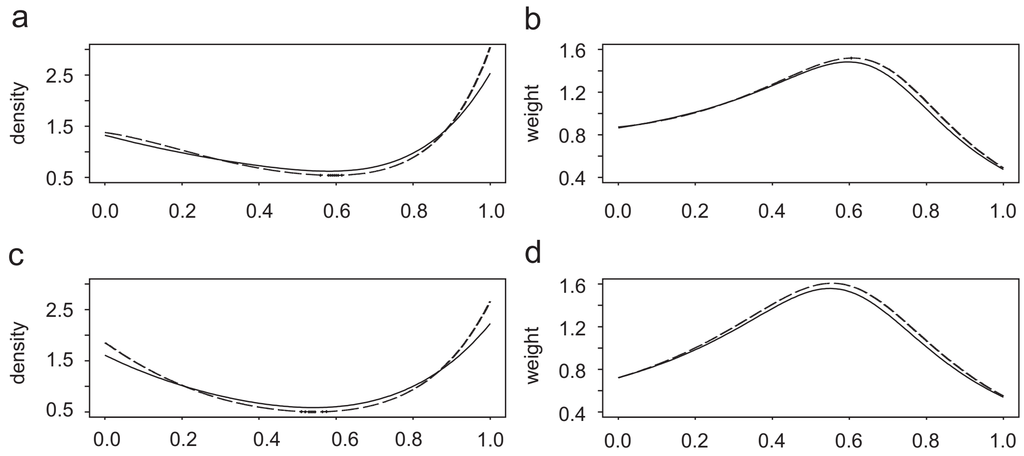

Fig. 3. Locally optimal and most robust design densities and corresponding weights for WLS in Example 6.3: (a) locally optimal design densities; (b) optimal weights corresponding to (a); (c) locally most robust design densities; and (d) optimal weights corresponding to (c). Each plot uses two values of v: $v = 0.5$ (solid line) and $v = 2$ (broken line).

Table 5
Coefficient values for the locally most robust product (23) of density and weights in Example 6.3

v	a_0	a_1	a_2	a_3	a_5	a_6	a_7	a_8	a_9
0.5	454.81	1703.64	−838.72	822.42	13.29	79.77	92972.48	2543.26	13839.74
1	439.79	3903.54	−2821.33	2281.76	0.198	81.49	255487.7	299.60	30479.58
2	−4938.22	7237.60	−6652.43	3692.79	−3.95	74.05	95311.10	3999.84	56657.32

Example 6.3. Recall Example 5.2 and model (9) with $S = [0, 1]$. The locally optimal product of density and weights for the prediction problem is given by (23). We take $a_4 = 1$. For $\theta_1 = 1$, the numerical values of the constants in (23) are given in Table 4. See Fig. 3(a) and (b) for plots of the locally optimal design densities and the corresponding optimal regression weights.

For $I = [0.5, 1.5]$, the locally most robust products of density and weights are provided in Table 5 for varying v. In each case, we found that the least favourable θ_1 within I is 0.5. See Fig. 3(c) and (d) for plots.

7. Concluding remarks

We have derived minimax prediction and extrapolation designs for misspecified generalized linear response models in the following three cases: (i) using OLS estimation under homoscedasticity, (ii) using OLS estimation under possible heteroscedasticity and (iii) using WLS estimation under possible heteroscedasticity. For each case with OLS, we conclude that the minimax extrapolation design density has the same form as that for the corresponding prediction problem. For case (iii), the product of the design density and weights function has the same form for both prediction and extrapolation. These analytic forms are completely general, but contain several constants to be determined numerically.

Wiens and Xu (2005) have derived minimax designs for extrapolation to a single point. Although the current work has assumed an extrapolation space with positive Lebesgue measure, the designs for one point extrapolation can be derived informally as limits of those in this article, as follows.

(1) The minimax one-point extrapolation design density for (i) above was shown in Wiens and Xu (2005) to have the form

$$m(\mathbf{x}) = \left[\frac{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\gamma} + d}{\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}\}^2}\right]^+.$$

This is the special case of form (7) with $\mathbf{P} = (\boldsymbol{\beta}\boldsymbol{\gamma}^{\mathrm{T}} + \boldsymbol{\gamma}\boldsymbol{\beta}^{\mathrm{T}})/2$ and $\mathbf{Q} = \boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}$.

(2) The minimax one-point extrapolation design density for (ii) above was shown to have the form

$$m(\mathbf{x}) = \left[\frac{\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}\}\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\gamma}\} + d}{\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}\}^2 + b\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}\}^4}\right]^+.$$

This is the special case of (13) with $\mathbf{P} = (\boldsymbol{\beta}\boldsymbol{\gamma}^{\mathrm{T}} + \boldsymbol{\gamma}\boldsymbol{\beta}^{\mathrm{T}})/2$, $\mathbf{Q} = \boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}$ and $\mathbf{U} = \sqrt{b}\mathbf{Q}$.

(3) The minimax product of design densities and weights for (iii) was shown to have the form

$$m(\mathbf{x}) = \left[\frac{(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\gamma}) + d - c}{(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})^2}\right]^+,$$

where c satisfies the cubic equation

$$c^3 + b(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\beta)^2 c = b(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\beta)^2[(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\beta)(\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\gamma) + d]$$

and $b > 0$. This is the special case of (23) with $\mathbf{P} = (\boldsymbol{\beta}\boldsymbol{\gamma}^{\mathrm{T}} + \boldsymbol{\gamma}\boldsymbol{\beta}^{\mathrm{T}})/2$, $\mathbf{Q} = \beta\beta^{\mathrm{T}}$ and $\mathbf{U} = \sqrt{b}\mathbf{Q}$.

Acknowledgements

The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada. We thank Professors Jane Ye and Julie Zhou, University of Victoria, and Dr. Peilin Shi, University of British Columbia, for helpful discussions. We also thank two anonymous referees, whose helpful comments have led to an improved article.

Appendix A. Derivations

Proof of Theorem 1. In what follows a prime $(\cdot)'$ denotes the Fréchet derivative of $(\cdot)$, $\partial(\cdot)$ is the Clarke generalized gradient of $(\cdot)$ and $N_{m(\mathbf{x})\geqslant 0}(m)$ is the normal cone of $\{m : m(\mathbf{x})\geqslant 0\}$, i.e.

$$N_{m(\mathbf{x})\geqslant 0}(m) = \left\{\omega(\mathbf{x}) : \int_S \omega(\mathbf{x})(m_1 - m)\,\mathrm{d}\mathbf{x} \leqslant 0 \quad \text{for any } m_1(\mathbf{x})\geqslant 0\right\}.$$

(See SYZ for basic definitions.)

Define $L_1 = \sup_{f\in\mathscr{F}} \mathrm{IMSEE}(f, \mathbf{1}, \mathbf{1}, m)$, given at (6), and let $m(\mathbf{x})$ be a density minimizing L_1. (The existence of such a density is established as in Ye and Zhou, 2004.) Then by the nonsmooth Lagrange multiplier rule (Clarke, 1983, Theorem 6.1.1), there exist real numbers $\lambda \geqslant 0$ and δ, not both zero, such that

$$0 \in \lambda\partial L_1(m) + \delta\left(\int_S m(\mathbf{x})\,\mathrm{d}\mathbf{x} - 1\right)' + N_{m(\mathbf{x})\geqslant 0}(m). \tag{A.1}$$

Note that $\mathbf{G} = \int_S [\{m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1}\}\tilde{\mathbf{z}}(\mathbf{x})][\{m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1}\}\tilde{\mathbf{z}}(\mathbf{x})]^{\mathrm{T}}\,\mathrm{d}\mathbf{x} \geqslant \mathbf{0}$. We temporarily assume that $\mathbf{G}$ is positive definite. Then as at Theorem 2 of SYZ, the generalized gradient of $\lambda_{\max}(\mathbf{G}\mathbf{H}_T)$ at m is

$$\partial\lambda_{\max}(\mathbf{G}\mathbf{H}_T) = \mathrm{co}\left\{\left(\frac{\mathbf{w}^{\mathrm{T}}\mathbf{H}_T\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{w}}\right)' : \mathbf{w} \in M(m)\right\},$$

where

$$M(m) = \left\{ \mathbf{w} : \frac{\mathbf{w}^{\mathrm{T}}\mathbf{H}_T\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{w}} = \max_{\|\mathbf{w}\|=1} \frac{\mathbf{w}^{\mathrm{T}}\mathbf{H}_T\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{w}} \right\}$$

and

$$\mathrm{co}\, A = \left\{ \sum \lambda_i a_i : \lambda_i \geqslant 0, \ \sum \lambda_i = 1, a_i \in A \right\}$$

is the convex hull of set A. From the Chain Rule (Clarke, 1983, Theorem 2.3.10),

$$\begin{aligned} \partial\left(\sqrt{\lambda_{\max}(\mathbf{GH}_T)} + r_{T,S}\right)^2 &= \left(1 + \frac{r_{T,S}}{\sqrt{\lambda_{\max}(\mathbf{GH}_T)}}\right) \partial\lambda_{\max}(\mathbf{GH}_T) \\ &= \left(1 + \frac{r_{T,S}}{\sqrt{\lambda_{\max}(\mathbf{GH}_T)}}\right) \mathrm{co}\left\{ \left(\frac{\mathbf{w}^{\mathrm{T}}\mathbf{H}_T\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{w}}\right)' : \mathbf{w} \in M(m) \right\}. \end{aligned} \tag{A.2}$$

We require the following Fréchet derivatives, which can be calculated as in SYZ:

$$(\mathrm{tr}[\mathbf{B}^{-1}\mathbf{A}_T])' = -\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}), \tag{A.3}$$

$$\left(\frac{\mathbf{w}^{\mathrm{T}}\mathbf{H}_T\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{G}^{-1}\mathbf{w}}\right)'_m = \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\tilde{\mathbf{M}}_{\mathbf{w}}\tilde{\mathbf{z}}(\mathbf{x}) + \{\tilde{\mathbf{b}}_{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{z}}(\mathbf{x})\}^2 m(\mathbf{x}). \tag{A.4}$$

In (A.4), $\mathbf{w}$ is any vector in $\mathbb{R}$ and $\tilde{\mathbf{M}}_{\mathbf{w}}$ is a $p \times p$ symmetric matrix, $\tilde{\mathbf{b}}_{\mathbf{w}}$ a $p \times 1$ vector whose specific values are not important to us.

By (A.1)–(A.4), we have that

$$\begin{aligned} 0 \in \lambda\eta_S^2 &\left[\left(1 + \frac{r_{T,S}}{\sqrt{\lambda_{\max}(\mathbf{GH}_T)}}\right) \mathrm{co}\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\tilde{\mathbf{M}}_{\mathbf{w}}\tilde{\mathbf{z}}(\mathbf{x}) + (\tilde{\mathbf{b}}_{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{z}}(\mathbf{x}))^2 m(\mathbf{x}) : \mathbf{w} \in M(m)\} - v\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}) \right] \\ &+ \delta + N_{m(\mathbf{x})\geqslant 0}(m). \end{aligned} \tag{A.5}$$

Let $\mathbf{M}_{\mathbf{w}} = (1 + r_{T,S}/\sqrt{\lambda_{\max}(\mathbf{GH}_T)})\tilde{\mathbf{M}}_{\mathbf{w}}$, $\mathbf{b}_{\mathbf{w}} = (1 + r_{T,S}/\sqrt{\lambda_{\max}(\mathbf{GH}_T)})^{1/2}\tilde{\mathbf{b}}_{\mathbf{w}}$ and note that $(\int_S m(\mathbf{x})\,\mathrm{d}\mathbf{x} - 1)'_m = 1$. Then (A.5) becomes

$$\begin{aligned} 0 \in \lambda\eta_S^2 &[\mathrm{co}\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{M}_{\mathbf{w}}\tilde{\mathbf{z}}(\mathbf{x}) + (\mathbf{b}_{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{z}}(\mathbf{x}))^2 m(\mathbf{x}) : \mathbf{w} \in M(m)\} - v\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x})] \\ &+ \delta + N_{m(\mathbf{x})\geqslant 0}(m). \end{aligned}$$

It can be shown, as in the proof of Theorem 1 of SYZ, that $\lambda \neq 0$.

By the definition of convex hull, there exists a positive integer N, nonnegative scalars $\lambda_1, \ldots, \lambda_N$ with $\lambda_1 + \cdots + \lambda_N = 1$, $\mathbf{w}_i \in M(m) \subset \mathbf{R}^p$ and $\varepsilon \in N_{m(\mathbf{x})\geqslant 0}(m)$ such that

$$\begin{aligned} 0 &= \lambda\eta_S^2 \left[\sum_{i=1}^N \lambda_i \{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{M}_{\mathbf{w}_i}\tilde{\mathbf{z}}(\mathbf{x}) + (\mathbf{b}_{\mathbf{w}_i}^{\mathrm{T}}\tilde{\mathbf{z}}(\mathbf{x}))^2 m(\mathbf{x})\} - v\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x}) \right] + \delta + \varepsilon \\ &= \lambda\eta_S^2 \sum_{i=1}^N \lambda_i (\mathbf{b}_{\mathbf{w}_i}^{\mathrm{T}}\tilde{\mathbf{z}}(\mathbf{x}))^2 m(\mathbf{x}) - \lambda\eta_S^2\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) \left[v\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1} - \sum_{i=1}^N \lambda_i \mathbf{M}_{\mathbf{w}_i} \right] \tilde{\mathbf{z}}(\mathbf{x}) + \delta + \varepsilon. \end{aligned}$$

Consequently, there exists a constant symmetric matrix $\mathbf{P} = \lambda\eta_S^2\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})[v\mathbf{B}^{-1}\mathbf{A}_T\mathbf{B}^{-1} - \sum_{i=1}^N \lambda_i \mathbf{M}_{\mathbf{w}_i}]$, a constant positive semi-definite matrix $\mathbf{Q} = \lambda\eta_S^2\sum_{i=1}^N \lambda_i \mathbf{b}_{\mathbf{w}_i}\mathbf{b}_{\mathbf{w}_i}^{\mathrm{T}}$ and a constant $d = -\delta$ such that

$$0 = \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{Q}\tilde{\mathbf{z}}(\mathbf{x})m(\mathbf{x}) - \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{P}\tilde{\mathbf{z}}(\mathbf{x}) - d + \varepsilon. \tag{A.6}$$

From Proposition 3 of SYZ we see that $\varepsilon = 0$ almost everywhere on $\{\mathbf{x} \in S : m(\mathbf{x}) > 0\}$ and hence

$$\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{Q}\tilde{\mathbf{z}}(\mathbf{x})m(\mathbf{x}) - \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{P}\tilde{\mathbf{z}}(\mathbf{x}) - d = 0$$

for all $\mathbf{x}$ such that $m(\mathbf{x})>0$. Since $\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{Q}\tilde{\mathbf{z}}(\mathbf{x})=\sum_{i=1}^{N}\lambda_i\{\mathbf{b}_{\mathbf{w}_i}^{\mathrm{T}}\tilde{\mathbf{z}}(\mathbf{x})\}^2>0$ for almost all $\mathbf{x}\in S$ we obtain

$$m(\mathbf{x})=\frac{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{P}\tilde{\mathbf{z}}(\mathbf{x})+d}{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{Q}\tilde{\mathbf{z}}(\mathbf{x})}$$

for almost all $\mathbf{x}\in S$ such that $m(\mathbf{x})>0$. For those $\mathbf{x}\in S$ such that $m(\mathbf{x})=0$ we apply Proposition 3 of SYZ again to infer that $\varepsilon\leqslant 0$ a.e. and hence by (A.6), $\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{P}\tilde{\mathbf{z}}(\mathbf{x})+d\leqslant 0$. Consequently,

$$\frac{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{P}\tilde{\mathbf{z}}(\mathbf{x})+d}{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{Q}\tilde{\mathbf{z}}(\mathbf{x})}\leqslant 0$$

for almost all $\mathbf{x}\in S$ such that $m(\mathbf{x})=0$, and (7) follows in the case that $\mathbf{G}$ is positive definite. This unnecessary assumption may now be dropped by arguing in the same manner as in the proof of Theorem 1 in SYZ. □

Proof of Theorem 2. We give the proof only for extrapolation, that for prediction being similar but simpler. Define $l_m(\mathbf{x})=\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_T\tilde{\mathbf{z}}(\mathbf{x})$. Then from (12) we seek a density $m(\cdot)$ minimizing

$$L_2\overset{\text{def.}}{=}\sup_{f\in\mathscr{F},g\in\mathscr{G}}\mathrm{IMSEE}(f,g,\mathbf{1},m)=\eta_S^2\left\{(\sqrt{\lambda_{\max}(\mathbf{G}\mathbf{H}_T)}+r_{T,S})^2+v\Omega^{-1/2}\left[\int_S\{l_m(\mathbf{x})m(\mathbf{x})\}^2\,d\mathbf{x}\right]^{1/2}\right\}.$$

We again initially assume $\mathbf{G}>0$. As in the preceding proof there exist real numbers $\lambda\geqslant 0$ and δ, not both zero, such that

$$0\in\lambda\partial L_2(m)+\delta\left(\int_S m(\mathbf{x})\,d\mathbf{x}-1\right)'+N_{m(\mathbf{x})\geqslant 0}(m), \tag{A.7}$$

where the last two terms $(\int_S m(\mathbf{x})\,d\mathbf{x}-1)'$ and $N_{m(\mathbf{x})\geqslant 0}(m)$ are the same as those in the proof of Theorem 1. Note that L_1 and L_2 differ only in their variance terms. Using

$$\left(\left[\int_S\{l_m(\mathbf{x})m(\mathbf{x})\}^2\,d\mathbf{x}\right]^{1/2}\right)'$$
$$=\left(\int_S\{l_m(\mathbf{x})m(\mathbf{x})\}^2\,d\mathbf{x}\right)^{-1/2}\left\{l_m^2(\mathbf{x})m(\mathbf{x})-2\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_T\left(\int_S l_m(\mathbf{x})m^2(\mathbf{x})\tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\,d\mathbf{x}\right)\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x})\right\}$$

in the evaluation of (A.7) we obtain

$$0\in\lambda\eta_S^2\left\{\begin{array}{l}\mathrm{co}\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{M}_{\mathbf{w}}\tilde{\mathbf{z}}(\mathbf{x})+(\mathbf{b}_{\mathbf{w}}^{\mathrm{T}}\tilde{\mathbf{z}}(\mathbf{x}))^2m(\mathbf{x}):\mathbf{w}\in M(m)\}+v\Omega^{-1/2}(\int_S\{l_m(\mathbf{x})m(\mathbf{x})\}^2\,d\mathbf{x})^{-1/2}\\ \quad\times\{l_m^2(\mathbf{x})m(\mathbf{x})-2\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_T(\int_S l_m(\mathbf{x})m^2(\mathbf{x})\tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\,d\mathbf{x})\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x})\}\end{array}\right\}$$
$$+\delta+N_{m(\mathbf{x})\geqslant 0}(m).$$

As in the proof of Theorem 1, $\lambda\neq 0$.

Employing the definition of convex hull we assert the existence of a positive integer N, nonnegative scalars $\lambda_1,\ldots,\lambda_N$ with $\lambda_1+\cdots+\lambda_N=1$, $\mathbf{w}_i\in M(m)\subset\mathbf{R}^p$ and $\varepsilon\in N_{m(\mathbf{x})\geqslant 0}(m)$ such that

$$0=\lambda\eta_S^2\left[\begin{array}{l}\sum_{i=1}^{N}\lambda_i\{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{M}_{\mathbf{w}_i}\tilde{\mathbf{z}}(\mathbf{x})+(\mathbf{b}_{\mathbf{w}_i}^{\mathrm{T}}\tilde{\mathbf{z}}(\mathbf{x}))^2m(\mathbf{x})\}+v\Omega^{-1/2}(\int_S\{l_m(\mathbf{x})m(\mathbf{x})\}^2\,d\mathbf{x})^{-1/2}\\ \quad\times\{l_m^2(\mathbf{x})m(\mathbf{x})-2\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\mathbf{H}_T(\int_S l_m(\mathbf{x})m^2(\mathbf{x})\tilde{\mathbf{z}}(\mathbf{x})\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\,d\mathbf{x})\mathbf{B}^{-1}\tilde{\mathbf{z}}(\mathbf{x})\}\end{array}\right]$$
$$+\delta+\varepsilon.$$

Consequently, there exists a symmetric matrix

$$\mathbf{P}=\eta_S^2\left[v\Omega^{-1/2}\left(\int_S\{l_m(\mathbf{x})m(\mathbf{x})\}^2\,d\mathbf{x}\right)^{-1/2}\{\mathbf{P}_1+\mathbf{P}_1^{\mathrm{T}}\}-\sum_{i=1}^{N}\lambda_i\mathbf{M}_{\mathbf{w}_i}\right]$$

with $\mathbf{P}_1 = \mathbf{H}_T \left(\int_S l_m(\mathbf{x}) m^2(\mathbf{x}) \tilde{\mathbf{z}}(\mathbf{x}) \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x})\, \mathrm{d}\mathbf{x}\right) \mathbf{B}^{-1}$, a positive semi-definite matrix

$$\mathbf{Q} = \eta_S^2 \sum_{i=1}^{N} \lambda_i \mathbf{b}_{\mathbf{w}_i} \mathbf{b}_{\mathbf{w}_i}^{\mathrm{T}},$$

a positive definite matrix

$$\mathbf{U} = \eta_S v^{1/2} \left(\Omega \int_S \{l_m(\mathbf{x}) m(\mathbf{x})\}^2 \, \mathrm{d}\mathbf{x}\right)^{-1/4} \mathbf{H}_T,$$

and scalars

$$\varepsilon = \frac{\varepsilon_0}{\lambda}, \quad d = -\frac{\delta}{\lambda}$$

such that

$$0 = [\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) \mathbf{Q} \tilde{\mathbf{z}}(\mathbf{x}) + \{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) \mathbf{U} \tilde{\mathbf{z}}(\mathbf{x})\}^2] m(\mathbf{x}) - \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) \mathbf{P} \tilde{\mathbf{z}}(\mathbf{x}) - d + \varepsilon. \tag{A.8}$$

The proof is now completed in a manner essentially identical to that of Theorem 1. □

Proof of Theorem 3. Again we give the proof only for extrapolation. In a manner very similar to that in the preceding two proofs we find that there exists a symmetric matrix **P**, a positive semi-definite matrix **Q**, a positive definite matrix **U** and a constant d such that on the set where $m(\mathbf{x}) > 0$,

$$\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) \mathbf{Q} \tilde{\mathbf{z}}(\mathbf{x}) m(\mathbf{x}) + \{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) \mathbf{U} \tilde{\mathbf{z}}(\mathbf{x})\}^{2/3} m^{1/3}(\mathbf{x}) - \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) \mathbf{P} \tilde{\mathbf{z}}(\mathbf{x}) - d = 0.$$

Therefore the minimizing $m(\mathbf{x})$ is a solution to

$$a(\mathbf{x}) m^{1/3}(\mathbf{x}) + b(\mathbf{x}) m(\mathbf{x}) - c(\mathbf{x}) = 0, \tag{A.9}$$

where $a(\mathbf{x}) = \{\tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) \mathbf{U} \tilde{\mathbf{z}}(\mathbf{x})\}^{2/3}$, $b(\mathbf{x}) = \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) \mathbf{Q} \tilde{\mathbf{z}}(\mathbf{x})$ and $c(\mathbf{x}) = \tilde{\mathbf{z}}^{\mathrm{T}}(\mathbf{x}) \mathbf{P} \tilde{\mathbf{z}}(\mathbf{x}) + d$. Let $\tilde{m} = c - bm$. Then, (A.9) becomes

$$\tilde{m}^3 + \frac{a^3}{b} \tilde{m} - \frac{a^3 c}{b} = 0. \tag{A.10}$$

Since a and b are positive almost everywhere in S, (A.10) has only one real solution. Applying Cardano's formula for cubic equations (Dunham, 1990), we obtain $\tilde{m}(\mathbf{x}) = k(\mathbf{x})$, where k is as at (24). Thus

$$m(\mathbf{x}) = \frac{c(\mathbf{x}) - k(\mathbf{x})}{b(\mathbf{x})}$$

on the set where $m(\mathbf{x}) > 0$. The rest of the proof is now essentially identical to that of Theorem 1. □

References

Adewale, A., 2002. Robust designs for approximate regression models with two interacting regressors. Unpublished M.Sc. Thesis, University of Alberta.

Atkinson, A.C., Haines, L.M., 1996. Designs for nonlinear and generalized linear models. In: Ghosh, S., Rao, C.R. (Eds.), Handbook of Statistics, vol. 13. pp. 437–475.

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

Clarke, F.H., 1983. Optimization and Nonsmooth Analysis. Wiley, New York.

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. In: Journey through Genius: The Great Theorems of Mathematics.Wiley, New York, pp. 133–154 (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.

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

Ford, I., Torsney, B., Wu, C.F.J., 1992. The use of a canonical form in the construction of locally optimal designs for non-linear problems. J. Roy. Statist. Soc. B 54, 569–583.

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

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.

Lawless, J.F., 1984. Some problems concerning experimental designs for extrapolation. In: Dwivedi, T.D. (Ed.), Topics in Applied Statistics. Concordia University Press, Montreal, pp. 357–366.

Shi, P., 2002. Minimax robust designs for misspecified regression models. Unpublished Ph.D. Thesis, University of Victoria.

Shi, P., Ye, J., Zhou, J., 2003. Minimax robust designs for misspecified regression models. Canad. J. Statist. 31, 397–414.

Silvey, S.D., 1980. Optimal Design. Chapman & Hall, London.

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.

Wiens, D.P., 1998. Minimax robust designs and weights for approximately specified regression models with heteroscedastic errors. J. Amer. Statist. Assoc. 93, 1440–1450.

Wiens, D.P., Xu, X., 2005. Robust designs for one-point extrapolation. University of Alberta Statistics Centre Technical Report 05.01. ⟨http://www.stat.ualberta.ca/stats_centre/tech.htm⟩.

Ye, J., Zhou, J., 2004. Existence and Symmetry of Minimax Regression Designs, preprint.