The Canadian Journal of Statistics
Vol. 27, No. 4, 1999, Pages 751–770
La revue canadienne de statistique

Robust extrapolation designs and weights for biased regression models with heteroscedastic errors

Zhide FANG and Douglas P. WIENS

Key words and phrases: Approximate linearity; bioassay; D-optimality; dose response; homoscedasticity; Q-optimality; prediction; weighted least squares.

AMS 1991 subject classifications: Primary 62K05, 62F35; secondary 62J05.

ABSTRACT

We consider the construction of designs for the extrapolation of regression responses, allowing both for possible heteroscedasticity in the errors and for imprecision in the specification of the response function. We find minimax designs and correspondingly optimal estimation weights in the context of the following problems: (1) for ordinary least squares estimation, determine a design to minimize the maximum value of the integrated mean squared prediction error (IMSPE), with the maximum being evaluated over both types of departure; (2) for weighted least squares estimation, determine both weights and a design to minimize the maximum IMSPE; (3) choose weights and design points to minimize the maximum IMSPE, subject to a side condition of unbiasedness. Solutions to (1) and (2) are given for multiple linear regression with no interactions, a spherical design space and an annular extrapolation space. For (3) the solution is given in complete generality; as one example we consider polynomial regression. Applications to a dose-response problem for bioassays are discussed. Numerical comparisons, including a simulation study, indicate that, as well as being easily implemented, the designs and weights for (3) perform as well as those for (1) and (2) and outperform some common competitors for moderate but undetectable amounts of model bias.

RÉSUMÉ

Les auteurs expliquent comment construire des plans d'expérience pour l'extrapolation de variables modélisées par régression en présence (i) d'hétéroscédasticité de l'erreur et (ii) d'imprécision dans la spécification du modèle. Les plans proposés sont minimax et les poids d'estimation correspondants sont optimaux dans les situations où : (1) on cherche un plan minimisant la valeur maximale sur (i) et (ii) de l'erreur quadratique moyenne de prévision intégrée (IMSPE) dans un contexte d'estimation par les moindres carrés ordinaires ; (2) on cherche à la fois un plan et des poids qui minimisent l'IMSPE maximal dans un contexte d'estimation par la méthode des moindres carrés pondérés ; (3) on veut sélectionner les points à échantillonner et les poids de façon à minimiser l'IMSPE maximal sous une condition d'absence de biais. Les solutions aux problèmes (1) et (2) sont données dans le cadre de la régression linéaire multiple sans interactions, pour un espace d'échantillonnage sphérique et pour un espace d'extrapolation annulaire. La solution au problème (3) est donnée en toute généralité et illustrée dans le cas de la régression polynomiale. Les auteurs présentent en outre des applications ayant trait à un problème de dose de réponse dans des bio-essais. Des comparaisons numériques prenant notamment la forme de simulations indiquent qu'en plus de leur facilité d'implantation, les plans et les poids optimaux du cas (3) se comportent aussi bien que ceux correspondant aux cas (1) et (2) en plus de surclasser certains compétiteurs d'usage courant dans des situations où le biais inhérent au modèle est relativement important sans toutefois être détectable.

1. INTRODUCTION

In this article we study the construction of designs for the extrapolation of regression responses, in the presence of both possible error heteroscedasticity and an approximately (and possibly incorrectly specified) response function. Design problems for *estimation* in the face of response uncertainty, but for *homoscedastic* errors, have been studied by Box & Draper (1959), Huber (1975), Pesotchinsky (1982), Wiens (1992) and others; Wiens (1998) allows also for heteroscedastic errors. Designs under error heteroscedasticity, assuming the fitted response to be exactly correct, were considered by Wong (1992) and Wong & Cook (1993); both of these papers assumed a *known* variance structure. Designs for extrapolation of polynomials, again assuming a correctly specified response, were studied by Kiefer & Wolfowitz (1964a, 1964b) and Hoel & Levine (1964). Studden (1971) studied such problems for multivariate polynomial models. Spruill (1984) and Dette & Wong (1996) constructed extrapolation designs for polynomial regression, robust against various misspecifications of the degree of the polynomial. Draper & Herzberg (1973) extended the methods of Box & Draper (1959) to extrapolation under response uncertainty. In their approach one estimates a first-order model but designs with the possibility of a second-order model in mind; the goal is extrapolation to one fixed point outside of the spherical design space. Huber (1975) obtained designs for extrapolation of a response (assumed to have a bounded derivative of a certain order but to be otherwise arbitrary) to one point outside the design interval. These results were corrected and extended by Huang & Studden (1988).

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. For such reasons we shall depart rather broadly from the usual linear model:

1. The response is taken to be only *approximately* linear in the regressors; *viz.*

$$\mathrm{E}(Y|\mathbf{x}) = \boldsymbol{\theta}'\mathbf{z}(\mathbf{x}) + f(\mathbf{x}) \tag{1}$$

   for a p-dimensional vector $\mathbf{z}$ of regressors, depending on a q-dimensional vector $\mathbf{x}$ of independent variables. The response contaminant f represents uncertainty about the exact nature of the regression response and is unknown and arbitrary, subject to certain restrictions detailed in Section 2. One estimates $\boldsymbol{\theta}$ but not f, leading possibly to biased estimation of $\mathrm{E}(Y|\mathbf{x})$ and consequently to biased predictions. The experimenter is to take n uncorrelated observations $Y_i = \mathrm{E}(Y|\mathbf{x}_i) + \varepsilon(\mathbf{x}_i)$, with $\mathbf{x}_i$ freely chosen from a design space S. The goal is to extrapolate the estimates of $\mathrm{E}(Y|\mathbf{x})$ to a given region T disjoint from S.

2. The random errors, although uncorrelated with mean zero, are possibly heteroscedastic: $\mathrm{var}\{\varepsilon(\mathbf{x})\} = \sigma^2 g(\mathbf{x})$ for a function g satisfying assumptions given in Section 2.

As an optimality criterion we take an analogue of the classical notion of Q-optimality: the supremum, over f and g, of the integrated mean squared prediction error (IMSPE) of the predicted values $\hat{Y}(\mathbf{x})$, with the integration being over the extrapolation region T, is to be minimized by an appropriate choice of design. The following problems will be addressed:

(P1) For *ordinary* least-squares (OLS) estimation, determine designs to minimize the maximum value, over f and g, of the IMSPE.

(P2) For *weighted* least-squares WLS estimation, determine designs and weights to minimize the maximum IMSPE.

(P3) Choose weights and design points to minimize the maximum IMSPE, subject to a side condition of unbiasedness.

As a possible application, consider the following extrapolation problem for bioassays. Let $P(x)$ be the probability of a particular response when a drug is administered at dose x. At various levels of x one observes the proportion p_x of subjects exhibiting the response and transforms to the p_x-quantile $Y = G^{-1}(p_x)$ for a suitable distribution such as the logistic. The regression function is then modelled as $\mathrm{E}(Y|x) \simeq G^{-1}(P(x))$. Since $P(x)$ is unknown, $\mathrm{E}(Y|x)$ is often approximated by a low-degree polynomial $\zeta(x)$. In the *low-dose* problem, it is difficult or impossible to observe Y near $x = 0$ or the error variance increases markedly as $x \to 0$; either of these situations leads to the extrapolation of estimates computed from data observed at, say, $x \in [a, 1]$ $(a > 0)$ to estimate $\mathrm{E}(Y|x = 0)$. Krewski, Bickis, Kovar & Arnold (1986) consider designs for such problems assuming that $\mathrm{E}(Y|x)$ is exactly linear in $\ln x$. Lawless (1984) takes an approach closer to ours, obtaining designs which minimize the MSPE of $\hat{Y}|_{x=0}$, for various trial values of $\mathrm{E}(Y|x=0)-\zeta(0)$. Of course this difference is unknown; the approach of the current article is to model it by $f(0)$ in (1) in such a way as to open the door to a minimax treatment. Another point of departure of our approach from that of Lawless (1984) or Huber (1975) is that although our treatment does not allow the case $T=\{0\}$ (or any other extrapolation space of Lebesgue measure zero), it *does* treat the case of an interval T, i.e., extrapolation to a *range* of values near $x=0$. This is particularly significant if the problem is to determine a "virtually safe dose" (Cornfield 1977).

Despite these differences, Lawless (1984) reaches qualitative conclusions very similar to ours, remarking that "... in extrapolation problems a slight degree of model inadequacy quickly wipes out advantages that minimum variance designs possess when the model is exactly correct".

The designs and weights which constitute solutions to problems (P1), (P2) and (P3) are given in Sections 3, 4 and 5, respectively. Those for (P1) and (P2) are theoretically and numerically rather complex, and our solutions are restricted to situations exhibiting considerable structure. In contrast, the solution to (P3) is given in complete generality and turns out to be computationally straightforward. We apply the solution to (P3) to the dose-response problem described above. A comparative study accompanied by concluding remarks and recommendations is given in Section 6. Some mathematical preliminaries are detailed in Chapter 2, where we reduce each of (P1)–(P3) to a single minimization over a class of densities. Proofs for Section 2 are postponed to the Appendix.

2. PRELIMINARIES AND NOTATION

For the regression model described in the Introduction, we shall assume that the contamination function $f(\mathbf{x})$ in (1) is an unknown member of the class

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

for positive constants η_S and η_T. The random errors $\varepsilon(\mathbf{x}_i)$ satisfy

$$\mathrm{var}\{\varepsilon(\mathbf{x}_i)\}=\sigma^2 g(\mathbf{x}_i),\qquad \int_S g^2(\mathbf{x})\,d\mathbf{x}\le\Omega^{-1}:=\int_S d\mathbf{x}<\infty.\tag{3}$$

The last condition of (2) is required in order that the true parameter $\boldsymbol{\theta}$ be uniquely defined, and then $\boldsymbol{\theta}:=\arg\min_{\mathbf{t}}\int_S\{\mathbf{t}'\mathbf{z}(\mathbf{x})-\mathrm{E}(Y|\mathbf{x})\}^2\,d\mathbf{x}$. One can instead *start* with this definition of $\boldsymbol{\theta}$, then *define* $f(\mathbf{x})=\boldsymbol{\theta}'\mathbf{z}(\mathbf{x})-\mathrm{E}(Y|\mathbf{x})$, thus obtaining the last condition of (2) as a natural consequence of the definition of the parameter being estimated. The other conditions of (2) are needed to ensure that errors due to estimation and prediction bias remain bounded. The conditions of (3) are equivalent to defining $\sigma^2=\sup_g[\int_S \mathrm{var}^2\{\varepsilon(\mathbf{x})\}\Omega\,d\mathbf{x}]^{1/2}$.

At the outset the only assumptions made about T are that it is disjoint from S and that the integrals in (2) exist; special cases will be considered in Sections 3 to 5. The requirement that T and S be disjoint need not exclude the application of our results to interpolation problems, i.e., the case $T\subset S$, as long as design points are not to be chosen from within T. One can then replace S by $S\backslash T$. The cases in which $S\subset T$, or in which S and T are merely overlapping, may be handled similarly. If design points may be chosen from within T, then $f(\mathbf{x})$ is defined for values $\mathbf{x}\in S\cap T$ and our method of maximizing the loss over $\mathcal{F}$ fails.

We remark that for (P1) and (P2) our results depend on the unknown parameters only through $\nu:=\sigma^2/(n\eta_S^2)$ and $r_{T,S}:=\eta_T/\eta_S$; for (P3) no knowledge whatsoever is required of these parameters. One can interpret ν as representing the relative importance of bias versus variance in the mind of the experimenter. As $\nu\to 0$ bias completely dominates the problem, whereas $\nu\to\infty$ results in a "pure variance" problem. Similarly, the choice of $r_{T,S}$ reflects the relative amounts of model response uncertainty in the extrapolation and design spaces. In our simulation study for this article, we made the rather arbitrary choice $\nu = \Omega$ and the intuitively appealing choice $r_{T,S} = 1$; perhaps an equally appealing choice of $r_{T,S}$ is the ratio of the volume of T to that of S. The qualitative aspects of the results did not change when other choices of ν and $r_{T,S}$ were made.

To avoid trivialities and 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 $\{\mathbf{x} \in S \cup T : \mathbf{a}'\mathbf{z}(\mathbf{x}) = \mathbf{0}\}$ has Lebesgue measure zero.

We propose to estimate $\boldsymbol{\theta}$ by least squares, possibly weighted with nonnegative weights $w(\mathbf{x})$. Let ξ be the design measure, i.e., $\xi = n^{-1}\Sigma_{i=1}^{n}\delta_{\mathbf{x}_i}$, where $\delta_{\mathbf{x}}$ is a point mass at $\mathbf{x}$. Define matrices and vectors

$$\begin{aligned}
\mathbf{A}_T &= \textstyle\int_T \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})\,d\mathbf{x}, & \mathbf{A}_S &= \textstyle\int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})\,d\mathbf{x},\\
\mathbf{B} &= \textstyle\int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})w(\mathbf{x})\,\xi(d\mathbf{x}), & \mathbf{D} &= \textstyle\int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})w^2(\mathbf{x})g(\mathbf{x})\,\xi(d\mathbf{x}),\\
\mathbf{b}_{f,S} &= \textstyle\int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})w(\mathbf{x})\,\xi(d\mathbf{x}), & \mathbf{b}_{f,T} &= \textstyle\int_T \mathbf{z}(\mathbf{x})f(\mathbf{x})\,d\mathbf{x}.
\end{aligned}$$

It follows from (A) that $\mathbf{A}_T$ and $\mathbf{A}_S$ are nonsingular and that $\mathbf{B}$ is nonsingular if ξ does not place mass on sets of Lebesgue measure zero. As discussed below, this sufficient requirement turns out to be necessary as well.

The WLS estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \mathbf{B}^{-1} \cdot \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}(\mathbf{x}_i)w(\mathbf{x}_i)Y_i = \mathbf{B}^{-1}\int_S \mathbf{z}(\mathbf{x})w(\mathbf{x})y(\mathbf{x})\,\xi(d\mathbf{x}),$$

with bias vector and covariance matrix

$$\mathrm{E}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta} = \mathbf{B}^{-1}\mathbf{b}_{f,S}, \qquad \mathrm{cov}(\hat{\boldsymbol{\theta}}) = \frac{\sigma^2}{n}\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}.$$

Note that $\sigma^2/n = \eta_S^2\nu$; we shall henceforth use the latter expression, since it will generally appear together with functions of the bias.

We predict $\mathrm{E}(Y|\mathbf{x})$ for $\mathbf{x} \in T$ by $\hat{Y}(\mathbf{x}) = \hat{\boldsymbol{\theta}}'\mathbf{z}(\mathbf{x})$ and consider the resulting IMSPE. The IMSPE splits into terms due to prediction bias, prediction variance and model misspecification:

$$\begin{aligned}
\mathrm{IMSPE}\,(f, g, w, \xi) &= \int_T \mathrm{E}[\{\hat{Y}(\mathbf{x}) - \mathrm{E}(Y|\mathbf{x})\}^2]\,d\mathbf{x}\\
&= \mathrm{IPB}\,(f, w, \xi) + \mathrm{IPV}\,(g, w, \xi) + \int_T f^2(\mathbf{x})\,d\mathbf{x},
\end{aligned}$$

where, with $\mathbf{H} := \mathbf{B}\mathbf{A}_T^{-1}\mathbf{B}$, the integrated prediction bias (IPB) and integrated prediction variance (IPV) are

$$\begin{aligned}
\mathrm{IPB}\,(f, w, \xi) &= \int_T [\mathrm{E}\{\hat{Y}(\mathbf{x}) - \boldsymbol{\theta}'\mathbf{z}(\mathbf{x})\}]^2\,d\mathbf{x} - 2\int_T \mathrm{E}\{\hat{Y}(\mathbf{x}) - \boldsymbol{\theta}'\mathbf{z}(\mathbf{x})\}f(\mathbf{x})\,d\mathbf{x}\\
&= \mathbf{b}_{f,S}'\mathbf{H}^{-1}\mathbf{b}_{f,S} - 2\mathbf{b}_{f,T}'\mathbf{B}^{-1}\mathbf{b}_{f,S}, \qquad (4)
\end{aligned}$$

$$\mathrm{IPV}\,(g, w, \xi) = \int_T \mathrm{var}\{\hat{Y}(\mathbf{x})\}\,d\mathbf{x} = \eta_S^2\nu\int_S \mathbf{z}'(\mathbf{x})\mathbf{H}^{-1}\mathbf{z}(\mathbf{x})w^2(\mathbf{x})g(\mathbf{x})\,\xi(d\mathbf{x}).$$

In contrast to the decomposition of IMSE for *estimation* into positive summands, the IPB may be negative. However, IPB$+ \int_T f^2(\mathbf{x})\,d\mathbf{x} \geq 0$.

In practice ξ must be discrete, with atoms consisting of integral multiples of n^{-1} at the design points. We adopt the viewpoint of *approximate* design theory and allow ξ to be any probability measure on S. It then turns out that the optimal extrapolation designs are not discrete. In fact, to guarantee that either of $\sup_f \text{IPB}(f, w, \xi)$ or $\sup_g \text{IPV}(g, w, \xi)$ will be finite, it is necessary that ξ have a density. This can be established by modifying the proof of Lemma 1 of Wiens (1992). A consequence is that the optimal extrapolation designs must be approximated to make them implementable. This can be carried out by placing the design points at an appropriate number of quantiles of ξ.

Let $k(\mathbf{x})$ be the density of ξ, and define $m(\mathbf{x}) = k(\mathbf{x})w(\mathbf{x})$. Without loss of generality, assume that the mean weight is $\int_S w(\mathbf{x})\,\xi(d\mathbf{x}) = 1$. Then $m(\mathbf{x})$ is also a density on S which for fixed weights satisfies

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

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

Given fixed $m(\mathbf{x})$ and $w(\mathbf{x})$, the "max" parts of the minimax solutions are given by Theorem 2.1. Before stating this, we define matrices $\mathbf{K} = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})m^2(\mathbf{x})\,d\mathbf{x}$ and $\mathbf{G} = \mathbf{K} - \mathbf{B}\mathbf{A}_S^{-1}\mathbf{B}$. We define λ_m to be the largest solution to $|\mathbf{G} - \lambda\mathbf{H}| = 0$ and let $\mathbf{a}_0$ be any vector satisfying $(\mathbf{G}\mathbf{H}^{-1}\mathbf{G} - \lambda_m\mathbf{G})\mathbf{a}_0 = \mathbf{0}$ and $\mathbf{a}_0'\mathbf{G}\mathbf{a}_0 = 1$. Define also $l_m(\mathbf{x}) = \mathbf{z}'(\mathbf{x})\mathbf{H}^{-1}\mathbf{z}(\mathbf{x})$ and $\alpha_m = \int_S \{l_m(\mathbf{x})m^2(\mathbf{x})\}^{2/3}\,d\mathbf{x}$.

THEOREM 2.1. (a) *The maximum integrated prediction bias is*

$$\sup_{f\in\mathcal{F}} \text{IPB}(f, w, \xi) = \eta_S^2\left\{\left(\sqrt{\lambda_m} + r_{T,S}\right)^2 - r_{T,S}^2\right\} \geq 0,$$

attained at

$$f_m(\mathbf{x}) = \begin{cases} \eta_S\mathbf{z}'(\mathbf{x})\{m(\mathbf{x})I - \mathbf{A}_S^{-1}\mathbf{B}\}\mathbf{a}_0, & \mathbf{x}\in S, \\ -\eta_T\mathbf{z}'(\mathbf{x})\mathbf{B}^{-1}\mathbf{G}\mathbf{a}_0/\sqrt{\lambda_m}, & \mathbf{x}\in T. \end{cases}$$

(b) *The maximum integrated prediction variance is*

$$\sup_g \text{IPV}(g, w, \xi) = \eta_S^2\nu\Omega^{-1/2}\left[\int_S \{w(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x})\}^2\,d\mathbf{x}\right]^{1/2},$$

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

(c) *The maximum integrated mean squared prediction error is*

$$\sup_{f,g} \text{IMSPE}(f, g, w, \xi) = \eta_S^2\left\{\left(\sqrt{\lambda_m} + r_{T,S}\right)^2 + \nu\Omega^{-1/2}\left[\int_S \{w(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x})\}^2\,d\mathbf{x}\right]^{1/2}\right\}.$$

Note that the least favourable contaminant is in fact *linear* (in $\mathbf{z}$) on T and that f_m also maximizes $\text{IPB} + \int_T f^2(\mathbf{x})\,d\mathbf{x}$ (since $\int_T f_m^2(\mathbf{x})\,d\mathbf{x} = \eta_T^2$).

We say that a design and weights pair (ξ, w) is *unbiased* if it satisfies

$$\text{E}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta} \qquad \text{for all } f \in \mathcal{F},$$

so that $\sup_f \text{IPB}(f, w, \xi) = 0$. The following theorem gives the minimax weights for fixed $m(\mathbf{x})$, and a necessary and sufficient condition for unbiasedness.

THEOREM 2.2. (a) *For fixed* $m(\mathbf{x})$ *the weights minimizing* $\sup_g$ IPV (g, w, ξ) *subject to* (5) *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 $\sup_g$ IPV $(g, w_m, \xi) = \eta_S^2 \nu \Omega^{-1/2} \alpha_m^{3/2}$.
(b) *The pair* (ξ, w) *is unbiased if and only if* $m(\mathbf{x}) \equiv \Omega$.

In view of Theorems 2.1 and 2.2, our problems in this article can be rewritten as follows:

(P1) Find a density $m_*(\mathbf{x})$ which minimizes

$$\eta_S^{-2} \sup_{f,g} \text{IMSPE}\,(f, g, \mathbf{1}, \xi) = \left(\sqrt{\lambda_m} + r_{T,S}\right)^2 + \nu \Omega^{-1/2} \left[\int_S \{l_m(\mathbf{x}) m(\mathbf{x})\}^2\, d\mathbf{x}\right]^{1/2}. \tag{6}$$

Then $k_*(\mathbf{x}) = m_*(\mathbf{x})$ is the optimal extrapolation design density for OLS estimation.

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

$$\eta_S^{-2} \sup_{f,g} \text{IMSPE}\,(f, g, w_m, \xi) = \left(\sqrt{\lambda_m} + r_{T,S}\right)^2 + \nu \Omega^{-1/2} \alpha_m^{3/2}.$$

Then the weights $w_*(\mathbf{x}) = m_*(\mathbf{x}) / k_*(\mathbf{x})$ and the design density

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

are optimal for WLS estimation.

(P3) Find weights $w_0(\mathbf{x}) \propto l_m(\mathbf{x})^{-2/3}$ satisfying (5) with $m(\mathbf{x}) \equiv \Omega$. Then the weights $w_0(\mathbf{x})$ and the design density $k_0(\mathbf{x}) = \Omega / w_0(\mathbf{x})$ are optimal in that they minimize $\sup_{f,g}$ IMSPE (f, g, w, ξ), subject to the side condition of unbiasedness.

Note that we have multiplied the quantities to be minimized by η_S^{-2}; this is without loss of generality and makes our results dependent only on the parameters ν and $r_{T,S}$.

3. MINIMAX EXTRAPOLATION DESIGNS FOR OLS

For (P1) and (P2) we consider only multiple linear regression without interactions, i.e., $\mathbf{z}'(\mathbf{x}) = (1, \mathbf{x}')$, with S being a q-dimensional sphere of unit radius centered at the origin. We take an annular extrapolation space: $T = \{\mathbf{x} | 1 < ||\mathbf{x}|| \leq \beta\}$. There being no reason to give preference to one coordinate of $\mathbf{x}$ over another, we restrict to densities $m(\mathbf{x})$ with identical, symmetric marginals. Then $\mathbf{A}_S = \Omega^{-1}(1 \oplus (q+2)^{-1} \mathbf{I}_q)$ where $\Omega = \Gamma(1 + q/2)/\pi^{q/2}$, and $\mathbf{B} = 1 \oplus \gamma \mathbf{I}_q$ where $\gamma := \int_S x_1^2 m(\mathbf{x})\, d\mathbf{x}$. Define parameters

$$\tau_0 = \int_T d\mathbf{x}, \qquad \kappa_0 = \Omega,$$
$$\tau_1 = \frac{\int_T x_1^2\, d\mathbf{x}}{q\gamma^2}, \qquad \kappa_1 = \Omega q (q+2) \gamma^2.$$

We calculate that

$$\int_T x_1^2\, d\mathbf{x} = \frac{\beta^{q+2} - 1}{\Omega(q+2)},$$

yielding $\tau_i = (\beta^{q+2i} - 1)/\kappa_i$ for $i = 0, 1$, and that $\mathbf{A}_T = \tau_0 \oplus \int_T x_1^2\, d\mathbf{x} \cdot \mathbf{I}_q$.

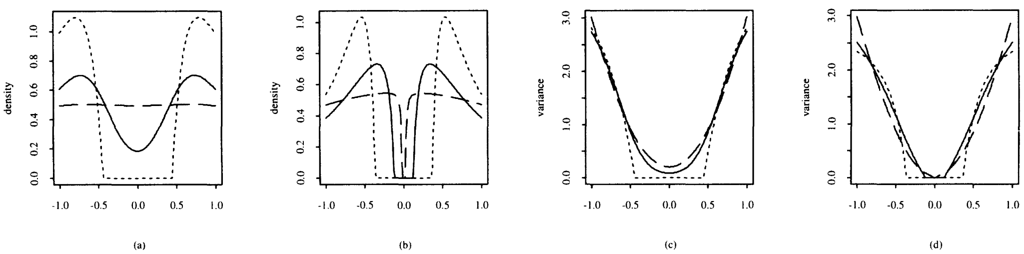

FIGURE 1: Optimal extrapolation design densities and least favourable variances for OLS and SLR: (a) design densities, $\beta = 1.5$; (b) design densities, $\beta = 5$; (c) least favourable variances, $\beta = 1.5$; (d) least favourable variances, $\beta = 5$. Each plot uses three values of ν: $\nu = 0.25$ (broken line), $\nu = 1$ (solid line), $\nu = 100$ (dotted line).

TABLE 1: Constants for $m_*(\mathbf{x})$ of Theorem 3.1; $q = 1$ (SLR) and $r_{T,S} = 1$ (* Case 0; † Case 1.).

β	ν	a	b	c	d	γ
1.5	0.25*	15.26	14.52	110.56	0.140	0.334
	0.5*	14.97	0.203	4.61	0.159	0.417
	1*	11.57	0.079	3.80	0.163	0.423
	10*	7.07	−0.163	1.53	0.184	0.579
	100*	6.80	−0.195	1.45	0.185	0.596
	∞*	6.77	−0.198	1.44	0.185	0.597
5	0.25†	7.01	−0.000	2.15	4.04e−6	0.334
	0.5†	2.51	−0.005	0.771	4.35e−6	0.339
	1†	1.48	−0.015	0.445	4.62e−6	0.343
	10†	0.983	−0.096	0.296	9.01e−6	0.436
	100†	0.981	−0.132	0.280	1.12e−5	0.475
	∞†	0.973	−0.134	0.277	1.13e−5	0.476

We find that $l_m(\mathbf{x}) = l(||\mathbf{x}||;\gamma)$, where $l(u;\gamma) := \tau_0 + \tau_1 q u^2$ depends on the design only through γ. The maximum eigenvalue λ_m in (6) is found to be $\lambda_m = \max(\lambda_m^{(0)}, \lambda_m^{(1)})$, where

$$\lambda_m^{(i)} := \tau_i \left\{ \int_S ||\mathbf{x}||^{2i} m^2(\mathbf{x})\, d\mathbf{x} - \kappa_i \right\}.$$

We must consider two cases. For $u \in [0,1]$ and $i = 0, 1$ define

$$h_i(u;\gamma) = \frac{a_i \nu (b_i + u^2)^+}{(1 + r_{T,S} c_i) u^{2i} + d_i \nu l^2(u;\gamma)},$$

where the constants $a_i = a_i(\gamma) > 0$, $b_i = b_i(\gamma)$, $c_i = c_i(\gamma) > 0$ and $d_i = d_i(\gamma) > 0$ satisfy

$$\int_0^1 \frac{qu^{q-1}}{\Omega} h_i(u;\gamma)du = 1, \tag{7}$$

$$\int_0^1 \frac{qu^{q-1}}{\Omega} u^2 h_i(u;\gamma)du = q\gamma, \tag{8}$$

$$c_i^2 \tau_i \left\{ \int_0^1 \frac{qu^{q-1}}{\Omega} u^{2i} h_i^2(u;\gamma)\, du - \kappa_i \right\} = 1, \tag{9}$$

$$2d_i\tau_i \left\{ \int_0^1 l^2(u;\gamma) qu^{q-1} h_i^2(u;\gamma)\, du \right\}^{1/2} = 1. \tag{10}$$

We denote these by Case 0 ($i = 0$) and Case 1 ($i = 1$). It turns out that for fixed ν, case 0 holds for small values of β, case 1 for large values of β. The precise relationship between ν and β has not been determined.

THEOREM 3.1 (Minimax extrapolation designs for OLS). *For $i = 0, 1$ define*

$$\gamma_i = \arg\min_{\gamma \geq 0} \left[\{c_i(\gamma)^{-1} + r_{T,S}\}^2 + \frac{\nu}{2\Omega d_i(\gamma)\tau_i} \right]. \tag{11}$$

If the inequality

$$\mathrm{E}_i\{\, U^{2(1-i)} h_i(U;\gamma_i)\} \leq \frac{c_i(\gamma_i)^{-2}}{\tau_{1-i}} + \kappa_{1-i} \tag{12}$$

holds, where $\mathrm{E}_i\{\,\cdot\,\}$ denotes expectation with respect to the density $(qu^{q-1}/\Omega)h_i(u;\gamma_i)$ and where κ_1 and τ_1 are evaluated at $\gamma = \gamma_1$, then the minimax (for OLS) extrapolation design density is

$$k_*(\mathbf{x}) = m_*(\mathbf{x}) = h_i(||\mathbf{x}||;\gamma_i).$$

The minimax IMSPE is

$$\sup_{f,g} \mathrm{IMSPE}\,(f, g, w = \mathbf{1}, \xi_*) = \eta_S^2 \left[\{c_i(\gamma_i)^{-1} + r_{T,S}\}^2 + \frac{\nu}{2\Omega d_i(\gamma_i)\tau_i} \right]. \tag{13}$$

Remarks.

(1) We sketch the proof of Theorem 3.1 for $i = 0$; that for $i = 1$ is similar. We first find m_0 minimizing (6) with $\lambda_m = \lambda_m^{(0)}$. Then if m_0 satisfies $\lambda_{m_0}^{(0)} \geq \lambda_{m_0}^{(1)}$, it is the required minimax density. For fixed γ and $\lambda_m = \lambda_m^{(0)}$, the loss (6) is a convex functional of m which remains fixed under orthogonal transformations of $\mathbf{x}$. By averaging over the orthogonal group we find that the minimizing m_0 is spherically symmetric. A standard variational argument shows that $m_0(\mathbf{x})$ is of the form $h_0(||\mathbf{x}||;\gamma)$ for appropriately chosen constants a_0–d_0. The integrand in (7) is the density of $U = ||\mathbf{x}||$, equation (8) fixes $\gamma = \mathrm{E}(U^2)/q$, equation (9) states that $c_0^{-2} = \lambda_m^{(0)}$, and equation (10) expresses the first-order variational condition that h_0 is a stationary point. These equations allow (6) to be expressed as a function of γ alone; a further minimization over γ then results in (11). The condition (12) ensures that $\lambda_{m_0}^{(0)}$ $(= c_0(\gamma_0)^{-2}) \geq \lambda_{m_0}^{(1)}$.

(2) For the numerical work, the equations (10) (for $i = 0, 1$) were first eliminated by using them in the presence of (7) to (9), to express d_i in terms of a_i–c_i:

$$d_i^{-1} = 4\Omega\tau_i^2 \left[a_i(b_i + q\gamma) - \frac{1 + r_{T,S}c_i}{\nu} \{(c_i^2\tau_i)^{-1} + \kappa_i\} \right]. \tag{14}$$

The remaining equations were then solved. Finally (11) was minimized and (12) verified. See Table 1 for some numerical values of the constants in the case $q = 1$ – straight-line regression (*SLR*) – with $r_{T,S} = 1$. Figure 1 gives plots of the minimax extrapolation design densities for varying β and ν.

(3) To implement these designs, we may use the fact that under the density $m_*(\mathbf{x})$, $\mathbf{x}/||\mathbf{x}||$ and $U = ||\mathbf{x}||$ are independently distributed, with $\mathbf{x}/||\mathbf{x}||$ being uniformly distributed over the surface of the unit sphere. A possible implementation is then as follows. Let H_* be the cumulative distribution function of U. Choose r_n design points uniformly distributed over each of the annuli $||\mathbf{x}|| = H_*^{-1}(i/[n/r_n])$, $i = 1, \ldots, [n/r_n]$, and $n - r_n(n/r_n)$ points at the origin.

(4) When the fitted model $\mathrm{E}(Y|\mathbf{x}) = \boldsymbol{\theta}'\mathbf{z}(\mathbf{x})$ is correct and the variances are homogeneous, the OLS estimate is unbiased and the loss is

$$\begin{aligned}\mathrm{IMSPE}\,(f = \mathbf{0},\ g = \mathbf{1},\ w = \mathbf{1},\ \xi) &= \mathrm{IPV}\,(\mathbf{1}, \mathbf{1}, \xi) = \eta_S^2 \nu \mathrm{E}\{l(||\mathbf{x}||; \gamma)\} \\ &= \eta_S^2 \nu \left(\tau_0 + \frac{q}{\gamma} \int_T x_1^2\, d\mathbf{x} \right),\end{aligned}$$

where γ is the second moment of ξ. In Figure 2(a) we compare the loss for our minimax *SLR* design ξ_* with that of the two-point (± 1) design ξ_1, constructed by Hoel & Levine (1964) under the assumption of an exactly correct fitted model and of the continuous uniform design ξ_2. When the model may contain response contamination and heteroscedastic errors, ξ_1 has $\sup_{f,g}$ IMSPE $= \infty$. Figure 2(b) gives plots of $\sup_{f,g}$ IMSPE for the uniform design and for ξ_*. For the minimax design $\sup_{f,g}$ IMSPE is given by (13). For the uniform design, Theorem 2.1(c) gives

$$\sup_{f,g} \mathrm{IMSPE}\,(f, g, \mathbf{1}, \xi_2) =$$

$$\eta_S^2 \left\{ r_{T,S}^2 + \nu \left[\{\tau_0 + q^2(q+2)\gamma^2\tau_1\}^2 + \frac{4q\{q(q+2)\gamma^2\tau_1\}^2}{q+4} \right]^{\frac{1}{2}} \right\}, \qquad (15)$$

with τ_1 evaluated at γ. We have used $\nu = \Omega = 0.5$ and $\eta_S = \eta_T = 1$ in Figure 2. For this value of ν the minimax design is close to the uniform, and the efficiencies relative to ξ_1, when the model is correct, are rather low. For larger values of ν these relative efficiencies are somewhat higher.

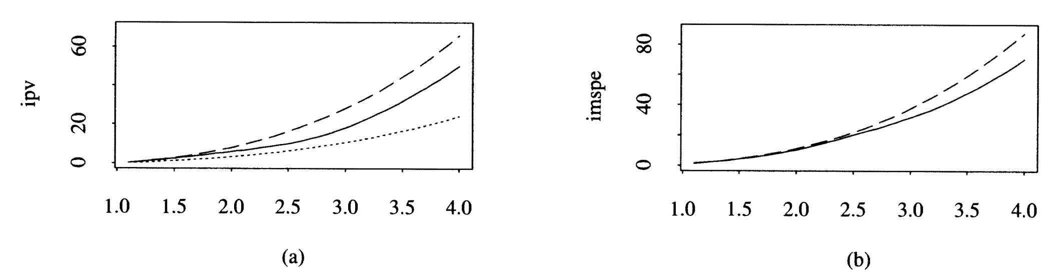

FIGURE 2: (a) Integrated prediction variance vs. β;
(b) $\sup_{f,g}$IMSPE vs. β for three designs: ξ_* (minimax, $\nu = \Omega$; solid lines), ξ_1 (two-point; dotted line), ξ_2 (uniform, broken lines); all for OLS and SLR.

4. MINIMAX EXTRAPOLATION DESIGNS AND WEIGHTS FOR WLS

We consider the same multiple linear regression model, spherical design space and extrapolation space as in the previous section. We again consider two cases. For $u \in [0,1]$ and $i = 0, 1$ define $h_i(u;\gamma)$ to be the (sole) real root of

$$\frac{1 + r_{T,S}c_i}{\nu}u^{2i}h_i(u;\gamma) + \left\{\frac{l^2(u;\gamma)h_i(u;\gamma)}{4\Omega^2 d_i \tau_i^4}\right\}^{1/3} - a_i(b_i + u^2)^+ = 0,$$

i.e., $h_i^{1/3}(u;\gamma) = z^{1/3}(u) - \left[\nu\{l^2(u;\gamma)/(4\Omega^2 d_i\tau_i^4)\}^{1/3}/\{3(1+r_{T,S}c_i)u^{2i}\}\right]z^{-1/3}(u)$, where

$$z(u) = \frac{\nu}{2(1 + r_{T,S}c_i)u^{2i}}\left\{a_i(b_i + u^2)^+ + \sqrt{\{a_i(b_i + u^2)^+\}^2 + \frac{\nu l^2(u;\gamma)}{27\Omega^2 d_i \tau_i^4(1 + r_{T,S}c_i)u^{2i}}}\right\}.$$

The constants $a_i = a_i(\gamma) > 0$, $b_i = b_i(\gamma)$, $c_i = c_i(\gamma) > 0$ and $d_i = d_i(\gamma) > 0$ are determined by (7), (8), (9) and (14).

The following result is established in a manner similar to that used for Theorem 3.1.

THEOREM 4.1 (Minimax extrapolation designs and weights for WLS). *For $i = 0, 1$ define*

$$\gamma_i = \arg\min_{\gamma \geq 0}\left[\{c_i(\gamma)^{-1} + r_{T,S}\}^2 + \frac{\nu}{4\Omega d_i(\gamma)\tau_i}\right].$$

If the inequality (12) holds, then the minimax (for WLS) extrapolation design density $k_*(\mathbf{x})$ and weights $w_{m_*}(\mathbf{x})$ are given by

$$\begin{aligned} k_*(\mathbf{x}) &= \left\{4\Omega^{1/2}d_i(\gamma_i)\tau_i h_i^2(||\mathbf{x}||;\gamma_i)l(||\mathbf{x}||;\gamma_i)\right\}^{2/3}, \\ m_*(\mathbf{x}) &= h_i(||\mathbf{x}||;\gamma_i), \qquad w_*(\mathbf{x}) = m_*(\mathbf{x})/k_*(\mathbf{x}). \end{aligned}$$

The minimax IMSPE *is*

$$\sup_{f,g} \text{IMSPE}\,(f, g, w_*, \xi_*) = \eta_S^2\left[\left\{c_i(\gamma_i)^{-1} + r_{T,S}\right\}^2 + \frac{\nu}{4\Omega d_i(\gamma_i)\tau_i}\right]. \tag{16}$$

Table 2 gives some typical values of the constants, and Figure 3 shows plots of the minimax design densities and weights—both for $q = 1$ and $r_{T,S} = 1$.

TABLE 2: Constants for $m_*(\mathbf{x})$ of Theorem 4.1; $q = 1$ (SLR) and $r_{T,S} = 1$. (*Case 0; † Case 1.)

β	ν	a	b	c	d	γ
1.5	0.25*	21.82	20.69	225.84	0.096	0.336
	0.5*	11.69	7.46	84.91	0.097	0.336
	1*	11.15	0.865	16.04	0.099	0.359
	10†	1.39	0.122	2.51	0.016	0.485
	100†	1.18	0.161	1.96	0.017	0.508
	∞†	1.16	0.166	1.91	0.017	0.511
5	0.25*	60.76	0.233	3.35	2.95e−5	0.349
	0.5†	3.44	0.005	0.875	5.26e−6	0.425
	1†	2.26	0.005	0.467	7.07e−6	0.485
	10†	1.38	0.027	0.247	9.68e−6	0.559
	100†	1.34	0.027	0.215	1.06e−5	0.585
	∞†	1.32	0.029	0.215	1.06e−5	0.585

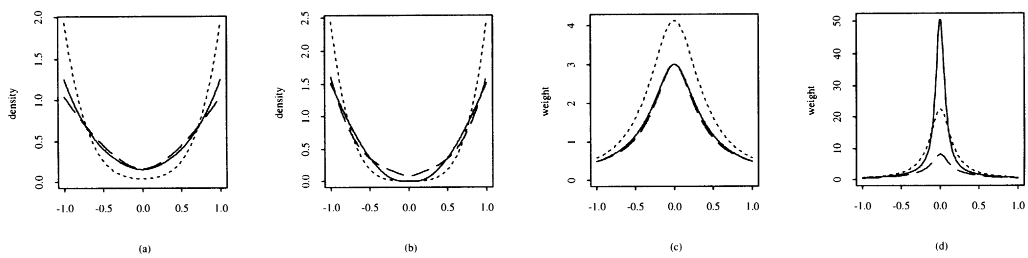

FIGURE 3: Optimal extrapolation design densities and minimax weights for WLS and SLR: (a) design densities, $\beta = 1.5$; (b) design densities, $\beta = 5$; (c) minimax weights, $\beta = 1.5$; (d) minimax weights, $\beta = 5$. Each plot uses three values of ν: $\nu = 0.25$ (broken line), $\nu = 1$ (solid line), $\nu = 200$ (dotted line).

We have computed the efficiencies of ξ_* relative to other designs ξ, also symmetric with identical marginals and with second moment γ. When the fitted model $\mathrm{E}(Y|\mathbf{x}) = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta}$ is correct and the variances are homogeneous, this relative efficiency is

$$\begin{aligned} \mathrm{re}\,1(\xi) &= \frac{\mathrm{IPV}\,(g = \mathbf{1},\ w = \mathbf{1}, \xi)}{\mathrm{IPV}\,(g = \mathbf{1},\ w = w_*, \xi_*)} \\ &= \frac{\tau_0 + (q/\gamma)\int_T x_1^2\,d\mathbf{x}}{(4\sqrt{\Omega}\,d_i\tau_i)^{-2/3}\int_0^1 (qu^{q-1}/\Omega)h_i^{2/3}(u;\gamma_i)l^{1/3}(u;\gamma_i)\,du}. \end{aligned}$$

Table 3 gives some representative values of $\mathrm{re}\,1(\xi)$ for $\xi = \xi_1$, with all mass on the boundary of S, and $\xi = \xi_2$, the continuous uniform design. Also given are values of

$$\mathrm{re}\,2(\xi) = \frac{\sup_{f,g} \mathrm{IMSPE}\,(f, g, \mathbf{1}, \xi)}{\sup_{f,g} \mathrm{IMSPE}\,(f, g, w_*, \xi_*)}, \tag{17}$$

which measures the efficiency of (ξ_*, w_*) relative to another design ξ, with constant weights, when the true response is only partially linear and the variances are heteroscedastic. The denominator of (17) is (16). For ξ_1 the numerator is ∞; for ξ_2 it is given by (15). As before, we take $\nu = \Omega$ and $r_{T,S} = 1$. The numbers in Table 3 show the appreciable gains to be enjoyed when ξ_* is employed in the presence of contamination and heteroscedasticity.

TABLE 3: Relative efficiencies re 1 (no contamination) and re 2 (maximal contamination) of ξ_* of Theorem 4.1, with optimal weights w_* and $\nu = \Omega$, versus the design ξ_1 with all mass on $||\mathbf{x}|| = 1$ and the uniform design ξ_2, both with constant weights. (* Case 0; † Case 1.)

β	q	$\mathrm{re}\,1(\xi_1)$	$\mathrm{re}\,1(\xi_2)$	$\mathrm{re}\,2(\xi_1)$	$\mathrm{re}\,2(\xi_2)$
1.5	1*	0.517	1.15	∞	1.27
	2*	0.609	1.07	∞	1.16
	3*	0.671	1.04	∞	1.10
5	1†	0.495	1.40	∞	1.57
	2†	0.567	1.11	∞	1.24
	3†	0.649	1.07	∞	1.14

5. OPTIMAL UNBIASED EXTRAPOLATION DESIGNS

In this section we make no *a priori* restrictions beyond assumption (A) on the design density, design space or extrapolation space. Note that if $m_0(\mathbf{x}) \equiv \Omega$, we have $\mathbf{B} = \Omega \mathbf{A}_S$. The following result is then an immediate consequence of Theorem 2.2.

THEOREM 5.1. *The density $k_0(\mathbf{x})$ of the optimal extrapolation design measure ξ_0, and optimal weights $w_0(\mathbf{x})$, which minimize* $\sup_{f,g}$ IMSPE (f, g, w, ξ) *subject to* $\sup_f$ IPB $(f, w, \xi) = 0$, *are given by*

$$k_0(\mathbf{x}) = \frac{\{\mathbf{z}'(\mathbf{x})\mathbf{A}_S^{-1}\mathbf{A}_T\mathbf{A}_S^{-1}\mathbf{z}(\mathbf{x})\}^{2/3}}{\int_S \{\mathbf{z}'(\mathbf{x})\mathbf{A}_S^{-1}\mathbf{A}_T\mathbf{A}_S^{-1}\mathbf{z}(\mathbf{x})\}^{2/3}\,d\mathbf{x}} \tag{18}$$

and $w_0(\mathbf{x}) = \Omega/k_0(\mathbf{x})$. The minimax IMSPE *is*

$$\sup_{f,g} \mathrm{IMSPE}\,(f, g, w_0, \xi_0) = \eta_S^2\left\{ r_{T,S}^2 + \nu\Omega^{-1/2}\left[\int_S \{\mathbf{z}'(\mathbf{x})\mathbf{A}_S^{-1}\mathbf{A}_T\mathbf{A}_S^{-1}\mathbf{z}(\mathbf{x})\}^{2/3}\,d\mathbf{x}\right]^{3/2}\right\},$$

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

Example 1. Consider the multiple linear regression model, design space and extrapolation space of Sections 3 and 4. The optimal unbiased extrapolation design density is

$$k_0(\mathbf{x}) \propto \left\{1 + (q+2)\frac{\beta^{q+2}-1}{\beta^q - 1}||\mathbf{x}||^2\right\}^{2/3}.$$

See Table 4 for the relative efficiencies, with $\nu = \Omega$ and $r_{T,S} = 1$. These efficiencies are at most only marginally lower than those of (ξ_*, w_*) of Section 4.

TABLE 4: Relative efficiencies re 1 (no contamination) and re 2 (maximal contamination) of ξ_0 of Example 1, with optimal weights w_0 and $\nu = \Omega$, versus the design ξ_1 with all mass on $||\mathbf{x}|| = 1$ and the uniform design ξ_2, both with constant weights.

β	q	re $1(\xi_1)$	re $1(\xi_2)$	re $2(\xi_1)$	re $2(\xi_2)$
1.5	1	0.514	1.14	∞	1.27
	2	0.607	1.07	∞	1.16
	3	0.671	1.04	∞	1.10
5	1	0.441	1.22	∞	1.54
	2	0.562	1.10	∞	1.23
	3	0.640	1.06	∞	1.13

Example 2. In this example there is insufficient structure to allow for a tractable treatment via (P1) or (P2), but (18) is easily evaluated. The regression response is as in Example 1, but the design space is the q-dimensional cube $S = [-1, 1]^q$ and the extrapolation region is the possibly asymmetric perimeter $T = [-\beta_1, \beta_2]^q \backslash S$, where $\beta_1, \beta_2 \geq 1$. One of β_1, β_2 may be unity, for one-sided extrapolation. We find that

$$k_0(\mathbf{x}) \propto \left\{\left(1 + 3\mu_1\Sigma_{i=1}^q x_i\right)^2 + 9\left(\mu_2 - \frac{1}{3\mu_3^q}\right)||\mathbf{x}||^2 - \frac{1}{\mu_3^q}\right\}^{2/3},$$

where $\mu_1 = (\beta_2 - \beta_1)/2$, $\mu_2 = (\beta_2 + \beta_1)^2/12$ and $\mu_3 = (\beta_2 + \beta_1)/2$. For symmetric extrapolation $\beta_1 = \beta_2$ and $\mu_1 = 0$.

For the dose-response problem discussed in Section 1, if a linear approximation to $E(Y|x)$ is taken then the design density is

$$\frac{2}{1-a}k_0\left\{\left(x-\frac{1+a}{2}\right)\left(\frac{2}{1-a}\right)\right\}, \quad a \le x \le 1;\ \mu_1 = \frac{a}{2},\ \mu_2 = \frac{a^2}{12}.$$

If instead a polynomial approximation is thought more appropriate, then the design density is obtainable by applying a similar linear transformation to x in Example 3 below. In either case, a suitable implementation would consist of taking an appropriate number of replicates at each of a number of quantiles of $\xi_0(\cdot)$. The number of replicates vs. the number of quantiles would likely be determined by the requirements of the particular problem under investigation.

Example 3: Polynomial regression. Take $\mathbf{z}'(x) = (1, x, \dots, x^{p-1})$, corresponding to polynomial regression of degree $p-1$, on $S = [-1, 1]$. To evaluate (18) it is convenient to first express $\mathbf{z}(x)$ in terms of the Legendre polynomials. Denote by $P_m(x)$ the mth-degree Legendre polynomial, normalized by $\int_{-1}^{1} P_m^2(x)\,dx = (m+0.5)^{-1}$. For instance $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = (3x^2-1)/2$, $P_3(x) = (5x^3-3x)/2$. We then find that

$$\mathbf{z}'(x)\mathbf{A}_S^{-1}\mathbf{A}_T\mathbf{A}_S^{-1}\mathbf{z}(x) = \sum_{0\le i,j\le p-1} \alpha_{ij}P_i(x)P_j(x),$$

where $\alpha_{ij} = (i+0.5)(j+0.5)\int_T P_i(x)P_j(x)\,dx$.

Denote the density (18) by $k_{p-1}(x;\beta)$. When T is symmetric, i.e., $T = [-\beta, \beta]\setminus S$, we find

$$\begin{aligned}
k_2(x;\beta) &\propto \{5\beta^3(\beta+1)(3x^2-1)^2 - \beta(\beta+1)(5x^4-22x^2+5) + 4(1-2x^2+5x^4)\}^{2/3},\\
k_3(x;\beta) &\propto \{175\beta^5(\beta+1)x^2(3-5x^2)^2 - 5\beta^3(\beta+1)(595x^6-963x^4+369x^2-9)\\
&\quad +5\beta(\beta+1)(140x^6-177x^4+90x^2-9) + 4(175x^6-165x^4+45x^2+9)\}^{2/3}.
\end{aligned}$$

When $T = [1, \beta]$ is one-sided, we find

$$\begin{aligned}
k_2(x;\beta) &\propto \{5\beta^4(3x^2-1)^2 + 5\beta^3(3x-1)(x+1)(3x^2-1)\\
&\quad -\beta^2(5x^4-30x^3-22x^2+10x+5)\\
&\quad -\beta(x+1)(5x^3-15x^2-7x+5) + 2(10x^4+5x^3-4x^2+x+2)\}^{2/3},\\
k_3(x;\beta) &\propto \{175\beta^6x^2(5x^2-3)^2 + 175\beta^5x(x+1)(5x^2-2x-1)(5x^2-3)\\
&\quad -5\beta^4(595x^6-525x^5-963x^4+490x^3+369x^2-105x-9)\\
&\quad +5\beta^3(x+1)(595x^5-385x^4-578x^3+258x^2+111x-9)\\
&\quad +5\beta^2(140x^6-210x^5-177x^4+320x^3+90x^2-102x-9)\\
&\quad +5\beta(x+1)(140x^5-35x^4-142x^3+48x^2+42x-9)\\
&\quad +700x^6+525x^5-660x^4-470x^3+180x^2+165x+36\}^{2/3}.
\end{aligned}$$

For both symmetric and one-sided extrapolation regions,

$$k_{p-1}(x;\infty) = \frac{\{P_{p-1}^2(x)\}^{2/3}}{\int_{-1}^{1}\{P_{p-1}^2(x)\}^{2/3}\,dx}. \tag{19}$$

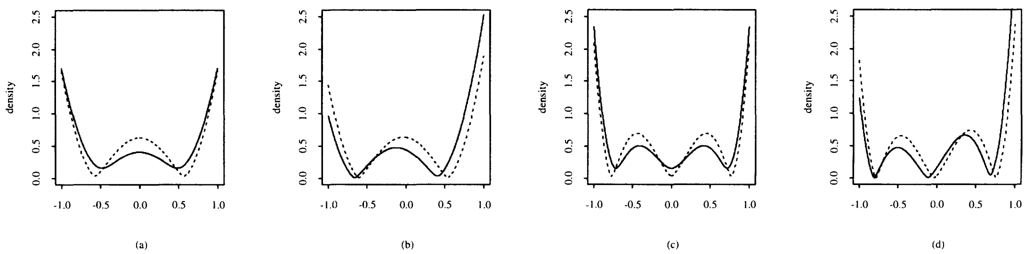

FIGURE 4:
Optimal unbiased extrapolation design densities in biased quadratic and cubic polynomial models: (a) quadratic model, symmetric extrapolation region; (b) quadratic model, one-sided extrapolation region; (c) cubic model, symmetric extrapolation region; (d) cubic model, one-sided extrapolation region. Each plot uses two values of β: $\beta = 1.5$ (solid line), $\beta = 5$ (dotted line).

Remarks.

(1) The limit in (19) is approached quite rapidly, and we find that for moderately large β the symmetric and one-sided design densities are, for practical purposes, identical. In contrast (see Figure 4), for small β and one-sided extrapolation the optimal designs place appreciably more mass on that side of S closer to the extrapolation region. For large p one can combine (19) with the asymptotic expansion

$$(p-1)P_{p-1}^2(x) = \frac{1}{\pi\sqrt{1-x^2}} + \frac{U_{2(p-1)}(x)}{\pi} + O(p^{-1/2}), \tag{20}$$

where $U_{2(p-1)}(x) = \sin((2p-1)\arccos x)/\sin(\arccos x)$ is Chebyshev's polynomial of the second kind. The right-hand side of (20) is a density whose first term is the limiting density of the D-optimal design as $p \to \infty$.

(2) The modes of $k_{p-1}(x;\infty)$ are at ± 1 and at the critical points of $P_{p-1}(x)$. Recall that these are precisely the support points of the classical D-optimal design which minimizes estimation variance alone. Thus $k_{p-1}(x;\infty)$ may be viewed as a smoothed version of the D-optimal design. Efficiencies relative to the continuous uniform design ξ_2 and Hoel & Levine's (1964) extrapolation design ξ_3, with $\nu = \Omega$ and $r_{T,S} = 1$, are given in Table 5. Note however that both ξ_3 and the D-optimal design have only as many design points as parameters, so that there is no opportunity to assess the fit of the model.

6. COMPARISONS

We have carried out a simulation study for a regression model as at (1)–(3) with $\mathbf{z}(x) = (1, x)'$ $(-1 \le x \le 1)$, normally distributed errors with $\sigma^2 = 1$ and sample size $n = 20$. We took $r_{T,S} = 1$ and $T = [-\beta, \beta]\backslash S$ with $\beta = 1.5$. Designs solving problems (P1), (P2) and (P3) were constructed and compared with the continuous uniform design ("U") and the two-point design ("HL") of Hoel & Levine (1964). Table 6 gives some values of $\eta_S^{-2}\max_{f,g}$ IMSPE. In preparing this table we assumed that (P2) and (P3) would be used with the correspondingly robust weights. Note that by this measure of maximum loss, the unbiased design (P3) performs as well as (P2) for moderate values of ν.

TABLE 5: Relative efficiencies re 1 (no contamination) and re 2 (maximal contamination) of ξ_0 of Example 3, with optimal weights w_0, versus the uniform design ξ_2 and the p-point design ξ_3, both unweighted.

		$T=[-\beta,\beta]\setminus S$				$T=(1,\beta]$			
β	$p-1$	re 1(ξ_2)	re 1(ξ_3)	re 2(ξ_2)	re 2(ξ_3)	re 1(ξ_2)	re 1(ξ_3)	re 2(ξ_2)	re 2(ξ_3)
1.5	1	1.14	.514	1.27	∞	1.35	.380	1.42	∞
	2	1.23	.387	1.78	∞	1.51	.329	2.27	∞
	3	1.32	.389	2.17	∞	1.53	.340	2.89	∞
5	1	1.25	.441	1.54	∞	1.30	.424	1.61	∞
	2	1.29	.455	1.76	∞	1.32	.452	1.83	∞
	3	1.29	.471	1.86	∞	1.32	.469	1.93	∞

TABLE 6: Comparative values of $\eta^{-2}\max_{f,g}$ IMSPE for the designs of Section 6.

ν	(P1)	(P2)	(P3)	(U)	(HL)
0.25	2.8	2.3	2.3	2.8	∞
0.5	4.6	3.6	3.6	4.6	∞
1	7.7	6.2	6.2	8.1	∞
10	57.1	48.4	53.1	72.5	∞
100	543	482	522	716	∞

To compare the relative performances against particular types of departures, we then chose a quadratic response: $f(x) \propto P_2(x)$ with the normalization $\int_S f^2(x)\,dx = 1/5$, and the variance function $g(x) \propto (1+x^2)^\alpha$ ($\alpha = 0, 2$), with the normalization $\int_S g^2(x)\,dx = \Omega^{-1}$. Designs (P1) and (P2) employed $\nu = \Omega = 0.5$. Note that then $\int_S f^2(x)\,dx = 2\eta_S^2$; this choice was made to further test the robustness of (P1) and (P2). For the continuous designs the design points were placed at the quantiles $\xi^{-1}((i-1)/(n-1))$ ($i = 1,\dots,n$) of the design measures. When using WLS, the weights used for (P1) were generated from Theorem 4.1 in the same way as those for (P2). The uniform design weights were generated from Theorem 5.1 in the same way as those for (P3). For (HL), with 10 points at each of ± 1, weighting has no effect. The other design points and weights were as shown in Table 7.

Table 8 gives values of η_S^{-2}IPB, η_S^{-2}IPV and η_S^{-2}IMSPE for both OLS and WLS fits. All three robust designs performed substantially better than did (U) or (HL); (P2) and (P3) in particular did well both with and without weights. Note however that when used without weights, this good performance was attained at the cost of a substantial negative IPB. When used with the optimal weights, (P2) and (P3) virtually eliminated this bias. Design (P3) enjoys the additional advantage of requiring no particular assumptions on the design space or fitted response function.

Faced with data reflecting the departures modeled by these simulations, would a statistician see evidence of the inadequacy of the linear model? To answer this we fitted a quadratic response $\theta_0 + \theta_1 x + \theta_2 x^2$ and carried out size-0.05 t-tests of $\mathcal{H}_0 : \theta_2 = 0$ vs. $\mathcal{H}_1 : \theta_2 \neq 0$. Both OLS and WLS fits were compared. The powers based on 20,000 simulations are presented in Table 9. The same 400,000 simulated normal errors were used in each of the four design cases. Note that for HL the quadratic model cannot be fitted and the power is zero.

TABLE 7: Design points and weights.

(P1)		(P2)		(P3)		(U)	
Design point	Weight	Design point	Weight	Design point	Weight	Design point	Weight
±0.095	2.49	±0.133	2.62	±0.148	2.59	±0.053	2.20
±0.265	1.71	±0.344	1.58	±0.353	1.57	±0.158	1.84
±0.398	1.24	±0.484	1.15	±0.489	1.15	±0.263	1.43
±0.506	0.976	±0.591	0.929	±0.595	0.934	±0.368	1.10
±0.600	0.811	±0.680	0.792	±0.682	0.800	±0.474	0.867
±0.684	0.698	±0.757	0.697	±0.759	0.705	±0.579	0.701
±0.764	0.612	±0.825	0.628	±0.827	0.636	±0.684	0.581
±0.842	0.544	±0.886	0.575	±0.889	0.583	±0.789	0.491
±0.920	0.487	±0.942	0.532	±0.947	0.539	±0.895	0.421
±1.00	0.438	±1.00	0.493	±1.00	0.504	±1.00	0.367

TABLE 8: IPB, IPV and IMSPE for the simulations of Section 6; heteroscedastic errors and contaminated response function. Values of IPV under homoscedasticity in parentheses.

	OLS			WLS		
Design	η^{-2}IPB	η^{-2}IPV	η^{-2}IMSPE	η^{-2}IPB	η^{-2}IPV	η^{-2}IMSPE
(P1)	$-.22$	0.30 (0.23)	1.09	.11	0.27 (0.27)	1.38
(P2)	$-.33$	0.28 (0.20)	0.96	$-.02$	0.25 (0.25)	1.23
(P3)	$-.33$	0.28 (0.20)	0.95	$-.04$	0.25 (0.24)	1.22
(U)	$-.07$	0.35 (0.26)	1.28	.29	0.30 (0.33)	1.59
(HL)	$-.86$	2.52 (1.29)	2.66	$-.86$	2.52 (1.29)	2.66

TABLE 9: Power of t-test of quadratic vs. linear response for homoscedastic ($\alpha = 0$) and heteroscedastic ($\alpha = 2$) errors.

	OLS		WLS	
Design	$\alpha = 0$	$\alpha = 2$	$\alpha = 0$	$\alpha = 2$
(P1)	0.29	0.31	0.29	0.33
(P2)	0.29	0.28	0.32	0.33
(P3)	0.29	0.27	0.32	0.32
(U)	0.31	0.37	0.20	0.29

A message to be learnt from the powers in Table 9 is that the common and realistic response departure used in these simulations is not likely to be detected, even when its parametric form is specified exactly by the alternative hypothesis. We view this as a powerful argument in favour of anticipating and addressing such departures at the design stage.

7. CONCLUSIONS AND GUIDELINES

We have given methods of designing for regression extrapolation, in the face of model uncertainties and possible heteroscedasticity, under some optimality criteria. The results tend to be somewhat complex and in some cases require extensive numerical work prior to implementation. They do however admit some informal and heuristic guidelines:

(1) In general, and as one would expect, the experimenter should place relatively more design points closer to the boundary between the design space S and the extrapolation space T, either as the volume of T increases relative to that of S, or as the emphasis on variance minimization versus bias minimization increases (as expressed by increasing values of ν).

(2) Notwithstanding the previous point, relative to designs for variance minimization alone the designs of this article are substantially more uniform, with mass spread throughout S rather than only at extreme points near T. This allows both for bias minimization and for the testing of alternative models.

(3) The unbiased designs of Section 5 are numerically less demanding than those of the preceding sections, although not completely without computational requirements. In line with (18), the general prescription is for the designer to place mass at points $\mathbf{x}$ proportional to values of $t(\mathbf{x}) := \{\mathbf{z}'(\mathbf{x})\mathbf{A}_S^{-1}\mathbf{A}_T\mathbf{A}_S^{-1}\mathbf{z}(\mathbf{x})\}^{2/3}$; the appropriate regression weights are then inversely proportional to this quantity. This requires a study of $t(\mathbf{x})$ for the particular design and extrapolation spaces under consideration. Some intuition can be gained from the explicit expressions in Examples 1 and 2; the latter in particular illustrates the manner in which the relative magnitude of $t(\mathbf{x})$ varies as T changes. As in Example 3, it can be convenient to transform to orthogonal regressors, so that $\mathbf{A}_S$ becomes a diagonal matrix.

A relevant problem concerns the manner in which the desired number n of observations is to be apportioned between design sites and replicates. We have recommended placing the former at quantiles of the optimal design densities; the determination of the number of such quantiles is the subject of further research.

APPENDIX: DERIVATIONS

Proof of Theorem 2.1. (a): First note that we can assume that the inequalities in (2) are in fact equalities. For, if $f \in \mathcal{F}$ is such that $\int_S f^2(\mathbf{x})\,d\mathbf{x} < \eta_S^2$ or $\int_T f^2(\mathbf{x})\,d\mathbf{x} < \eta_T^2$, then we define a function $cf \in \mathcal{F}$ as being $c_S f$ on S and $c_T f$ on T, where $|c_S| \geq 1$, $|c_T| \geq 1$ and the sign of $c_T c_S$ is chosen so that $-2\mathbf{b}'_{cf,T}\mathbf{B}^{-1}\mathbf{b}_{cf,S} = -2c_T c_S \mathbf{b}'_{f,T}\mathbf{B}^{-1}\mathbf{b}_{f,S} \geq -2\mathbf{b}'_{f,T}B^{-1}\mathbf{b}_{f,S}$. Then $\text{IPB}(cf,\xi) \geq \text{IPB}(f,\xi)$. Hence it is sufficient to evaluate the maximum value of $\text{IPB}(f,\xi)$ under the conditions $\int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})\,d\mathbf{x} = \mathbf{0}$, $\int_S f^2(\mathbf{x})\,d\mathbf{x} = \eta_S^2$, $\int_T f^2(\mathbf{x})\,d\mathbf{x} = \eta_T^2$.

Note that

$$\mathbf{G} = \int_S [\{m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1}\}\mathbf{z}(\mathbf{x})][\{m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1}\}\mathbf{z}(\mathbf{x})]'\,dx \geq \mathbf{0}. \tag{A.1}$$

We temporarily assume that $\mathbf{G}$ is positive definite. Given any $f \in \mathcal{F}$, define

$$h_f(x) = \begin{cases} s_f \mathbf{z}'(\mathbf{x})\{m(\mathbf{x})\mathbf{I} - \mathbf{A}_S^{-1}\mathbf{B}\}\mathbf{H}^{-1}\mathbf{b}_{f,S}, & x \in S, \\ t_f \mathbf{z}'(\mathbf{x})\mathbf{B}^{-1}\mathbf{b}_{f,S}, & x \in T, \end{cases}$$

with

$$s_f^2 = \eta_S^2/\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}, \quad t_f^2 = \eta_T^2/\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{b}_{f,S}$$

and $s_f = \pm\sqrt{s_f^2}$, $t_f = \pm\sqrt{t_f^2}$ chosen so that $\mathbf{b}'_{h_f,T}\mathbf{B}^{-1}\mathbf{b}_{h_f,S} \le 0$. Then we claim that (i) $h_f(x) \in \mathcal{F}$ and (ii) $\mathrm{IPB}(h_f,\xi) \ge \mathrm{IPB}(f,\xi)$. The verification of (i) is straightforward. For (ii) we note that $\mathbf{b}_{h_f,T} = t_f\mathbf{A}_T\mathbf{B}^{-1}\mathbf{b}_{f,S}$, that $\mathbf{b}_{h_f,S} = s_f\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}$, and that

$$\int_S f(\mathbf{x})h_f(\mathbf{x})\,d\mathbf{x} = s_f\mathbf{b}'_{f,S}\mathbf{H}_{f,S}^{-1}\mathbf{b}_{f,S}, \qquad \int_T f(\mathbf{x})h_f(\mathbf{x})\,d\mathbf{x} = t_f\mathbf{b}'_{f,S}\mathbf{B}^{-1}\mathbf{b}_{f,T}. \qquad (\mathrm{A}.2)$$

Evaluating (4) gives

$$\mathrm{IPB}(h_f,\xi) = s_f^2\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S} + 2|s_f|\,|t_f|\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}.$$

By the first equality of (A.2) and the Cauchy-Schwarz inequality, we have $s_f^2(\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{b}_{f,S})^2 \le \int_S f^2(\mathbf{x})\,d\mathbf{x}\int_S h_f^2(\mathbf{x})\,d\mathbf{x} \le \eta_S^4$, so that the definition of s_f gives $\eta_S^2 \ge (\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{b}_{f,S})^2/\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}$. Similarly, $\eta_T^2 \ge (\mathbf{b}'_{f,S}\mathbf{B}^{-1}\mathbf{b}_{f,T})^2/\mathbf{b}_{f,S}\mathbf{H}^{-1}\mathbf{b}_{f,S}$. Hence

$$\begin{aligned}
\mathrm{IPB}\,(h_f,\xi) &= \frac{\eta_S^2(\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S})}{\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}} + \frac{2\eta_S\eta_T(\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S})}{\sqrt{(\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S})(\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{b}_{f,S})}}\\
&\ge \frac{(\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{b}_{f,S})^2}{(\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S})^2}\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S} + 2|\mathbf{b}'_{f,S}\mathbf{B}^{-1}\mathbf{b}_{f,T}|,
\end{aligned}$$

and so $\mathrm{IPB}(h_f,\xi) \ge \mathrm{IPB}(f,\xi)$ if

$$(\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{b}_{f,S})(\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S}) \ge (\mathbf{b}'_{f,S}\mathbf{H}^{-1}\mathbf{G}\mathbf{H}^{-1}\mathbf{b}_{f,S})^2,$$

an inequality whose verification is again straightforward.

We can now restrict to $f \in \mathcal{F}$ of the same form as h_f, i.e.,

$$f(\mathbf{x};\mathbf{a},\mathbf{c}) = \begin{cases} \mathbf{z}'(\mathbf{x})\{m(\mathbf{x})\mathbf{I} - \mathbf{A}_S^{-1}\mathbf{B}\}\mathbf{a}, & x \in S,\\ \mathbf{z}'(\mathbf{x})\mathbf{B}^{-1}\mathbf{c}, & x \in T, \end{cases}$$

where $\mathbf{a}$ and $\mathbf{c}$ satisfy $\eta_S^2 = \int_S f^2(\mathbf{x};\mathbf{a},\mathbf{c})\,dx = \mathbf{a}'\mathbf{G}\mathbf{a}$, $\eta_T^2 = \int_T f^2(\mathbf{x};\mathbf{a},\mathbf{c})\,dx = \mathbf{c}'\mathbf{H}^{-1}\mathbf{c}$. Subject to these conditions we are to maximize $\mathrm{IPB}\,(f,\xi) = \mathbf{a}'\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{a} - 2\mathbf{c}'\mathbf{H}^{-1}\mathbf{G}\mathbf{a}$. The maximizing $\mathbf{c}$ is $\mathbf{c} = -\eta_T\mathbf{G}\mathbf{a}/||\mathbf{H}^{-1/2}\mathbf{G}\mathbf{a}||$, and then $\mathrm{IPB}\,(f,\xi) = (\sqrt{\mathbf{a}'\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{a}} + \eta_T)^2 - \eta_T^2$. With $\mathbf{a}_0 = \mathbf{a}/\eta_S$, we are then to maximize $\mathbf{a}_0'\mathbf{G}\mathbf{H}^{-1}\mathbf{G}\mathbf{a}_0$ subject to $\mathbf{a}_0'\mathbf{G}\mathbf{a}_0 = 1$. This is a standard eigenvalue problem. If λ_m is the largest solution to $|\mathbf{G}\mathbf{H}^{-1}\mathbf{G} - \lambda\mathbf{G}| = 0$, i.e., $|\mathbf{G} - \lambda\mathbf{H}| = 0$, then the maximizing $\mathbf{a}_0$ is a solution to $(\mathbf{G}\mathbf{H}^{-1}\mathbf{G} - \lambda_m\mathbf{G})\mathbf{a}_0 = \mathbf{0}$, normalized to satisfy $\mathbf{a}_0'\mathbf{G}\mathbf{a}_0 = 1$. A final evaluation of $\mathrm{IPB}(f,\xi)$ now completes the proof of (a) when $\mathbf{G} > \mathbf{0}$.

If the design density $m(\mathbf{x})$ is such that $\mathbf{G} = \mathbf{G}(m) \ge \mathbf{0}$ but $|\mathbf{G}| = 0$, we proceed as follows. Take any density $m_1(\mathbf{x})$ for which the corresponding matrix $\mathbf{G}(m_1) > \mathbf{0}$. Put $m_t(\mathbf{x}) = (1 - t)m(\mathbf{x}) + tm_1(\mathbf{x})$ and define $p(t) = |\mathbf{G}(m_t)|$. Then $p(t)$ is a polynomial in $t \in [0,1]$ with $p(0) = 0$ and $p(1) > 0$, so that $p(t)$ is nonconstant and nonnegative on $[0,1]$. Thus $p(t) > 0$ for all sufficiently small $t > 0$. Now apply (a) of the theorem to $\mathbf{G}(m_t)$, and let $t \to 0$, to see that the result holds in the general case.

(b): By the Cauchy-Schwarz inequality we have

$$\int_S w(\mathbf{x})g(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x})\,d\mathbf{x} \le \left[\int_S \{w(\mathbf{x})l_m(\mathbf{x})m(\mathbf{x})\}^2\,d\mathbf{x}\right]^{1/2}\left\{\int_S g^2(\mathbf{x})\,d\mathbf{x}\right\}^{1/2}$$

and (b) follows. Part (c) follows from (a) and (b). □

Proof of Theorem 2.2. Part (a) is a straightforward variational problem. For (b), note that by Theorem 2.1(a) and (A.1) we have

$$\sup_f \mathrm{IPB}\,(f, w, \xi) = 0 \iff \lambda_m = 0 \iff \mathbf{G} = \mathbf{0}$$
$$\iff (m(\mathbf{x})\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1})\mathbf{z}(\mathbf{x}) = 0 \text{ a.e.}$$

Thus $m(\mathbf{x})$ is an eigenvalue of $\mathbf{B}\mathbf{A}_S^{-1}$ if $\mathbf{z}(\mathbf{x}) \neq \mathbf{0}$, so that on $S_0 := \{\mathbf{x} \in S : \mathbf{z}(\mathbf{x}) \neq \mathbf{0}\}$, $m(\mathbf{x})$ can assume at most p distinct values. Decompose S_0 as $S_0 = \bigcup_{i=1}^{s} S_i$, with $s \leq p$ and $m(\mathbf{x}) \equiv \alpha_i$ on S_i. For any S_i with positive Lebesgue measure the relationship $(\alpha_i\mathbf{I} - \mathbf{B}\mathbf{A}_S^{-1})\mathbf{z}(\mathbf{x}) \equiv \mathbf{0}$, together with assumption (A), forces $\alpha_i\mathbf{I} = \mathbf{B}\mathbf{A}_S^{-1}$, so that at most one set S_i can have positive measure. Thus $m(\mathbf{x})$ is almost everywhere constant on S_0, hence on S itself since, again by (A), $S \setminus S_0$ is of measure zero. □

ACKNOWLEDGEMENTS

We are grateful for the thorough review by an anonymous referee of a previous version of this paper, and for the helpful comments of this referee and the Associate Editor and Editor. This research is supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

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

J. Cornfield (1977). Carcinogenic risk assessment. *Science*, 198, 693–699.

H. Dette & W. K. Wong (1996). Robust optimal extrapolation designs. *Biometrika*, 83, 667–680.

N. R. Draper & A. Herzberg (1973). Some designs for extrapolation outside a sphere. *Journal of the Royal Statistical Society Series B*, 35, 268–276.

P. G. Hoel & A. Levine (1964). Optimal spacing and weighting in polynomial prediction. *The Annals of Mathematical Statistics*, 35, 1533–1560.

M. N. L. Huang & W. J. Studden (1988). Model robust extrapolation designs. *Journal of Statistical Planning and Inference*, 18, 1–24.

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

J. Kiefer & J. Wolfowitz (1964a). Optimum extrapolation designs, I. *Annals of the Institute of Statistical Mathematics*, 16, 79–108.

J. Kiefer & J. Wolfowitz (1964b). Optimum extrapolation designs, II. *Annals of the Institute of Statistical Mathematics*, 16, 295–303.

D. Krewski, M. Bickis, J. Kovar & D. L. Arnold (1986). Optimal experimental designs for low dose extrapolation. I. The case of zero background. *Utilitas Mathematica*, 29, 245–262.

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

L. Pesotchinsky (1982). Optimal robust designs: linear regression in R . *The Annals of Statistics*, 10, 511–525.

M. C. Spruill (1984). Optimal designs for minimax extrapolation. *Journal of Multivariate Analysis*, 15, 52–62.

W. J. Studden (1971). Optimal designs for multivariate polynomial extrapolation. *The Annals of Mathematical Statistics*, 42, 828–832.

D. P. Wiens (1992). Minimax designs for approximately linear regression. *Journal of Statistical Planning and Inference*, 31, 353–371.

D. P. Wiens (1998). Minimax robust designs and weights for approximately specified regression models with heteroscedastic errors. *Journal of the American Statistical Association*, 93, 1440–1450.

W.-K. Wong (1992). A unified approach to the construction of minimax designs. *Biometrika*, 79, 611–619.

W.-K. Wong & R. D. Cook (1993). Heteroscedastic G-optimal designs. *Journal of the Royal Statistical Society Series B*, 55, 871–880.

Received 3 January 1998
Accepted 25 September 1998

Zhide FANG: zfang@math.uno.edu
Dept. of Mathematics, University of New Orleans
Lakefront, New Orleans, LA 70148-0001, USA

Douglas P. WIENS: doug.wiens@ualberta.ca
Statistics Centre, Dept. of Mathematical Sciences
University of Alberta, Edmonton, Alberta, Canada T6G 2G1